Designing Embodied Interactive Software Agents for E-Learning: Principles, Components, and Roles

Aktualisierte, gekürzte und korrigierte Fassung

INAUGURAL-DISSERTATION

zur Erlangung des Grades eines
Doktors der Philosophie (Dr. phil.)

dem Fachbereich Fremdsprachliche Philologien
der Philipps-Universität Marburg

vorgelegt von

Peter Franke, Magister Artium (M.A.)
geboren in Marburg/Lahn

Marburg/Lahn, 2009

Referenten: Prof. Dr. Jürgen Handke
Prof. Dr. Manfred Sommer
Prof. Dr. Ulrich Hasenkamp

Eingereicht am: 25. Januar 2008
Als Dissertation angenommen am: 23. September 2008
Tag der Disputation am: 24. Oktober 2008

Wissenschaftlicher Werdegang

Peter Franke hat an der Philipps-Universität Marburg Englische Sprachwissenschaft, Betriebswirtschaftslehre (mit Schwerpunkt Wirtschaftsinformatik) und Informatik studiert. Nach seinem Examen 1999 (Abschluss: „Magister Artium“, Prädikat: „Mit Auszeichnung“) arbeitete er zunächst von 1999 bis 2001 als wissenschaftliche Hilfskraft mit Abschluss, dann von 2001 bis 2008 als wissenschaftlicher Mitarbeiter von Prof. Dr. Jürgen Handke in der Abteilung Englische Sprachwissenschaft des Instituts für Anglistik und Amerikanistik an der Universität Marburg. Seit 2008 ist er als wissenschaftlicher Mitarbeiter im Projekt „Entwicklung von E-Learning-Angeboten für die Lehreraus- und Weiterbildung“, gefördert durch die Medieninitiative „Schule @ Zukunft“ des Landes Hessen, für Herrn Prof. Handke tätig.

Er gehört dem „Linguistic Engineering Team (LET)“ an, das seit 2001 die drei E-Learning-Umgebungen „The Virtual Linguistics Campus“ (www.linguistics-online.com), „The Virtual Center for Teacher Training“ (vct.online.uni-marburg.de) und den „LET-Vokabeltrainer“ (voctrainer.online.uni-marburg.de) entwickelt hat. Im Linguistic Engineering Team ist Peter Franke sowohl für die Entwicklung linguistischer und computerlinguistischer E-Learning-Inhalte als auch für die Konzeption und Implementierung von webbasierten Anwendungen zuständig.

Im Rahmen seiner Lehrtätigkeit bietet Peter Franke virtuelle und ‚blended‘ Lehrveranstaltungen in den Bereichen Theoretische Englische Sprachwissenschaft und Sprachtechnologien an. Seine Forschungsschwerpunkte sind verkörperte interaktive pädagogische Software-Agenten, Sprachtechnologien und E-Learning.

Academic Career

Peter Franke studied English linguistics, business administration (with an emphasis on business informatics), and computer science at Phillips University Marburg. After receiving his Master of Arts (Magister Artium) degree with Highest Distinction in 1999, he worked for Prof. Dr. Handke as a postgraduate student assistant from 1999 until 2001 and as a research assistant from 2001 until 2008 in the English linguistics division of the Department of English and American Studies at Philipps University Marburg. Since 2008 he has been working for Prof. Dr. Handke as a research assistant on the project "Development of E-Learning Courses for Basic and Continuing Teacher Education", which is funded by the German state of Hesse's media initiative "Schule @ Zukunft" (School @ Future).

Peter Franke is a member of the "Linguistic Engineering Team" (LET), which has developed three e-learning environments since 2001: "The Virtual Linguistics Campus" (www.linguistics-online.com), "The Virtual Center for Teacher Training" (www.vct.online.uni-marburg.de), and "The LET Vocabulary Trainer" (www.voctrainer.online.uni-marburg.de). As a member of the LET, he is responsible for the development of e-learning content in linguistics and human language technologies, as well as the design and implementation of web-based applications.

Peter Franke teaches both virtual and 'blended' courses in the fields of theoretical English linguistics and human language technologies. His research focuses on embodied interactive pedagogical software agents, human language technologies, and e-learning.

To My Parents.

Contents

List of Figures

List of Tables

List of Abbreviations

ACL	Agent Communication Language
ACT	Adaptive Control of Thought
AI	Artificial Intelligence
ANN	Artificial Neural Network
APA	Animated Pedagogical Agent
ASM	Abstract Shape Model
ASR	Automatic Speech Recognition
ATN	Augmented Transition Network
AU	Action Unit
AV-ASR	Audio-Visual Speech Recognition
BDI	Belief-Desire-Intention
CAI	Computer-Assisted Instruction
CATLM	Cognitive-Affective Theory of Learning with Media
CBM	Constraint-Based Modeling
CBR	Case-Based Reasoning
CBSM	Case-Based Student Modeling
CBT	Computer-Based Training
CD	Conceptual Dependency
CFG	Context-Free Grammar
CGI	Computer-Generated Imagery
CLI	Command-Line Interface
CMC	Computer-Mediated Communication
CMS	Content Management System
CTML	Cognitive Theory of Multimedia Learning
CTS	Concept-To-Speech
DAG	Directed Acyclic Graph
DAI	Distributed Artificial Intelligence
DL	Description Logics
DOF	Degree Of Freedom
DV	Dependent Variable
ECA	Embodied Conversational Agent

FACS	Facial Action Coding System
FAQ	Frequently Asked Question(s)
FD	Functional Description
FSA	Finite-State Automaton
FST	Finite-State Transducer
FSTN	Finite-State Transition Network
FTA	Face-Threatening Act
GPS	Global Positioning System
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HLT	Human Language Technologies
HMM	Hidden Markov Model
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICALL	Intelligent Computer-Assisted Language Learning
ICT	Information and Communication Technologies
ID	Instructional Design
IQ	Intelligence Quotient
ISO	International Organization for Standardization
iSTART	interactive Strategy Trainer for Active Reading and Thinking
IT	Information Technology
ITS	Intelligent Tutoring System
IV	Independent Variable
KIF	Knowledge Interchange Format
KQML	Knowledge Query and Manipulation Language
KRL	Knowledge Representation Language
LCMS	Learning Content Management System
LET	Linguistic Engineering Team
LMS	Learning Management System
LP	Learning Platform
LSS	Learning Support System
LVCSR	Large Vocabulary Continuous Speech Recognition
MAS	Multi-Agent System

MEA	Means-Ends Analysis
MLE	Managed Learning Environment
MMORPG	Massively Multiplayer Online Role-Playing Games
MPI	Measures of Perceived Interactivity
MUD	Multi-User Domain/Dungeon/Dimension
NL	Natural Language
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
NPC	Non-Player Character
OOP	Object-Oriented Programming
OWL	Web Ontology Language
PAL	Pedagogical Agent as Learning Companion
PCFG	Probabilistic Context-Free Grammar
PHP	PHP: Hypertext Preprocessor
PUI	Perceptual User Interface
RDF	Resource Description Framework
RDFS	RDF Schema
ROI	Region Of Interest
RSS	Rich Site Summary/Really Simple Syndication
RST	Rhetorical Structure Theory
RuleML	Rule Markup Language
SAM	Script-Applier Mechanism
SD	Standard Deviation
SWRL	Semantic Web Rules Language
TLTS	Tactical Language Training System
TMS	Truth Maintenance System
TTS	Text-To-Speech
UAS	User-Adaptive System
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VCT	Virtual Center for Teacher Training

VLC	Virtual Linguistics Campus
VLE	Virtual Learning Environment
VoIP	Voice over IP
VR	Virtual Reality
WIMP	Window, Icon, Menu, Pointer
XML	eXtensible Markup Language

Abstract

Embodied interactive software agents are complex autonomous, adaptive, and social software systems with a digital embodiment that enables them to act on and react to other entities (users, objects, and other agents) in their environment through bodily actions, which include the use of verbal and non-verbal communicative behaviors in face-to-face interactions with the user. These agents have been developed for various roles in different application domains, in which they perform tasks that have been assigned to them by their developers or delegated to them by their users or by other agents. In computer-assisted learning, embodied interactive pedagogical software agents have the general task to promote human learning by working with students (and other agents) in computer-based learning environments, among them e-learning platforms based on Internet technologies, such as the Virtual Linguistics Campus (www.linguistics-online.com). In these environments, pedagogical agents provide contextualized, qualified, personalized, and timely assistance, cooperation, instruction, motivation, and services for both individual learners and groups of learners.

This thesis develops a comprehensive, multidisciplinary, and user-oriented view of the design of embodied interactive pedagogical software agents, which integrates theoretical and practical insights from various academic and other fields. The research intends to contribute to the scientific understanding of issues, methods, theories, and technologies that are involved in the design, implementation, and evaluation of embodied interactive software agents for different roles in e-learning and other areas. For developers, the thesis provides sixteen basic principles (Added Value, Perceptible Qualities, Balanced Design, Coherence, Consistency, Completeness, Comprehensibility, Individuality, Variability, Communicative Ability, Modularity, Teamwork, Participatory Design, Role Awareness, Cultural Awareness, and Relationship Building) plus a large number of specific guidelines for the design of embodied interactive software agents and their components. Furthermore, it offers critical reviews of theories, concepts, approaches, and technologies from different areas and disciplines that are relevant to agent design. Finally, it discusses three pedagogical agent roles (virtual native speaker, coach, and peer) in the scenario of the linguistic fieldwork classes on the Virtual Linguistics Campus and presents detailed considerations for the design of an agent for one of these roles (the virtual native speaker).

Zusammenfassung

Verkörperte interaktive Software-Agenten sind komplexe autonome, adaptive und soziale Software-Systeme, die mit einem digitalen Körper ausgestattet sind, der sie in die Lage versetzt, auf andere Entitäten (Benutzer, Objekte und andere Agenten) in ihrer Umgebung durch körperliche Aktionen einzuwirken bzw. auf sie zu reagieren, was die Verwendung von verbalem und non-verbalem kommunikativem Verhalten in persönlichen Interaktionen „von Angesicht zu Angesicht" mit dem Benutzer einschließt. Diese Agenten sind für diverse Rollen in unterschiedlichen Anwendungsgebieten entwickelt worden, in denen sie Aufgaben erfüllen, die ihnen von ihren Entwicklern zugedacht bzw. von ihren Benutzern oder anderen Software-Agenten an sie delegiert worden sind. Im computerunterstützten Lernen haben verkörperte interaktive pädagogische Software-Agenten die generelle Aufgabe, menschliches Lernen zu fördern, indem sie mit Lernenden (und anderen Agenten) in computerbasierten Lernumgebungen arbeiten, wozu auch E-Learning-Plattformen auf der Basis von Internet-Technologien, wie z. B. der Virtual Linguistics Campus (www.linguistics-online.com), gehören. In diesen Umgebungen bieten pädagogische Agenten kontextualisierte, qualifizierte, personalisierte und zeitgerechte Hilfestellungen, Zusammenarbeit, Unterricht, Motivation und Dienstleistungen sowohl für einzelne Lernende als auch für Gruppen von Lernenden an.

Die vorliegende Arbeit entwickelt eine umfassende, multidisziplinäre und benutzerorientierte Sicht auf das Design von verkörperten interaktiven pädagogischen Software-Agenten, die theoretische und praktische Erkenntnisse aus verschiedenen akademischen und anderen Gebieten integriert. Sie will zum wissenschaftlichen Verständnis von Problemen, Methoden, Theorien und Technologien beitragen, die in das Design, die Implementierung und die Evaluierung von verkörperten interaktiven Software-Agenten für unterschiedliche Rollen im E-Learning und in anderen Gebieten involviert sind. Für Entwickler zeigt die Arbeit sechzehn grundlegende Prinzipien (Mehrwert, Erkennbare Qualitäten, Ausgeglichenes Design, Kohärenz, Konsistenz, Vollständigkeit, Verständlichkeit, Individualität, Variabilität, Kommunikationsfähigkeit, Modularität, Teamarbeit, Partizipatives Design, Rollenbewusstsein, Kulturelles Bewusstsein und Beziehungsaufbau) sowie eine große Zahl von speziellen Richtlinien für das Design von verkörperten interaktiven Software-Agenten und ihren Komponenten auf. Außerdem bietet sie kritische Bewertungen von Theorien, Konzepten, Ansätzen und Technologien aus unterschiedlichen Gebieten und Disziplinen, die für das Agentendesign wichtig sind. Schließlich diskutiert sie drei Rollen für pädagogische Agenten (Virtueller Muttersprachler, Trainer und Co-Lerner) im Szenario der linguistischen Feldarbeitskurse im Virtual Linguistics Campus und präsentiert detaillierte Überlegungen zum Design eines Agenten für eine dieser Rollen (den Virtuellen Muttersprachler).

Acknowledgements

The process of doing research for and writing up a Ph.D. thesis is a long journey, and there are times when the road becomes rough and steep, or forks into several directions without the convenience of a road sign nearby. At these points, it is a good thing to have others who cheer or spur you on (as the need may be), show you the way when you are lost, and provide many other kinds of support that facilitate your journey. Clearly, the present work would not have been possible without the help of many individuals over the years. Therefore, I gladly take this opportunity at the beginning to gratefully acknowledge their various contributions.

First, I would like to thank my advisor, Prof. Dr. Jürgen Handke, for bringing me into his team years ago and allowing me to play my part in building the Virtual Linguistics Campus (VLC), the e-learning environment that provided the background for my research on embodied interactive pedagogical software agents that is reported in this thesis. I am also grateful for the many things that I have learned during my time under his guidance, both from him as a mentor and from working for him on various projects. He encouraged and helped me to tackle and overcome quite a few challenges that I would not likely have tried my hand at otherwise. In particular during the final stages of the thesis project, his comments and support were invaluable.

Next, my sincere and heartfelt thanks go to all members of the Linguistic Engineering Team (LET) at the English department of Philipps University Marburg: Lili Dong, Anna Maria Gallus, Eva Großkurth, Anika Hampel, Marco Kurzrock, Sara J. Nere, Dirk Schäfer, Steffen Schaub, Daniel Schmidt, Alexander Sperl, Dilyana Toneva, Annett Vömel, Dirk Weidmann, and Songül Yokaribas. Furthermore, I would like to thank the following former team members and other contributors: Holger Becker, Wiebke Begere, Dennis Bock, Jens Bödicker, Uwe Colberg, Alexander Dröge, Catrina Engelbertz, Toni Gunner, Astrid Hente-Eickhorst, Dirk Hovy, Dr. Frauke Intemann, Matthias Katerbow, Jan Königsfeld, Elizabeth A. Kornmann, Ute Kreibaum, Reiko Oba, Mark Rothenpieler, Dennis Schwarz, Katja Stiegler, Tobias Unger, and Juliane Weiß. It has been/still is a pleasure and a privilege to work with all of you guys!

Special thanks go to my colleague Alexander Sperl for his assistance with the preparation of the figures in the thesis. I also gratefully acknowledge the outstanding artwork of Astrid Hente-Eickhorst, which can be seen in various items and screenshots taken from the VLC.

At home, I have to thank my family, in particular my parents to whom I have dedicated this thesis, for their unwavering support during the many long days and nights which I spent at my desk typing away on my laptop computer. I could not have done this without you. Thank you, Mom. Thank you, Dad.

Finally, above everyone else, I humbly thank my Lord and Savior Jesus Christ for being the one who has given me the strength and guidance I needed to go on every time I was ready to give up during this project. Without him, I can do nothing, but “I can do all things through Christ which strengtheneth me” (The Holy Bible, Philippians 4:13). It is my wish to honor him with my life, and I pray that this thesis is a small contribution to that.

—Peter Franke, M.A.
Bad Endbach-Günterod,
Hesse, Germany
September 2009

1 Introduction

For the average person in the modern, post-industrial world, especially for those who were born or grew up in it, a life without computers and other digital artifacts[1] created by the information technology (IT) industry is unimaginable. Many people use these artifacts on an everyday basis to work, play, learn, socialize, bond, break up, make love, war, and peace, and recently even to build alternate identities and economic existences in synthetic worlds (Castronova 2006). In short, ever greater parts of our daily lives are lived with, through, and dependent on computers. As Johnson et al. put it, "computing technology is becoming ubiquitous and ingrained in our society, often becoming the focal point of our businesses and social interactions" (Johnson et al. 2006: 447). It is certainly not an exaggeration to say that economies, governments, relationships, and individual lives thrive with, but may also crumble without, the help of computers.

1.1 Roles of Computers

In the present work, the term *computer* is taken to refer to a (typically electronic) device that can accept information (*data*) in a particular (usually digital) form (*input*) and can perform a sequence of operations on this data (*processing*), following a preset but variable sequence of instructions (*program*), to produce a result (*output*) in the form of information or signals. Computers may be stand-alone machines, but today, they are commonly connected with other computers or devices, typically over a *network*, such as the Internet, which serves as a medium for the exchange of messages generated by people or computers for interpretation by humans or machines.

The traditional strength of computers is in their capability to process large amounts of data in a short period of time. While raw processing power is still important, computers nowadays serve people in many ways other than just "crunching numbers." From the perspective of the *user*,[2] the various functions of computers can be classified in terms of three large categories (cf. Figure 1). Computers can serve people as tools, as media, and as social actors (Fogg 2003: Chapters 2–5):

- *Computers as tools*. The computer increases the capabilities of users. It helps them to accomplish tasks more easily or more efficiently, guides them through processes involving multiple steps, and does things for them they possibly could not do without technology.
- *Computers as media*. The computer provides an experience for users, by conveying symbolic information (text, graphics, charts, icons, etc.), sensory information (mostly audio and video, rarely smell and touch), or a combination of both; by giving users the opportunity to interact with the information or the environment; and by immersing users in computer-based simulations (cf. Chapter 2.3.4) that allow them to explore cause-and-effect relations, rehearse behaviors, and become involved in motivating vicarious experiences.

---

[1] An artifact is any product of human craft or any object modified by humans, e.g. a tool, vehicle, household appliance, mobile device, web site, or software system ([INT 1]).

[2] A user is a person who makes use of a thing (here: the computer) to achieve something, for himself or herself or for other individuals, alone or in cooperation or competition with other users, for work, learning, or pleasure. Characteristics of users are discussed in Chapter 6.1.

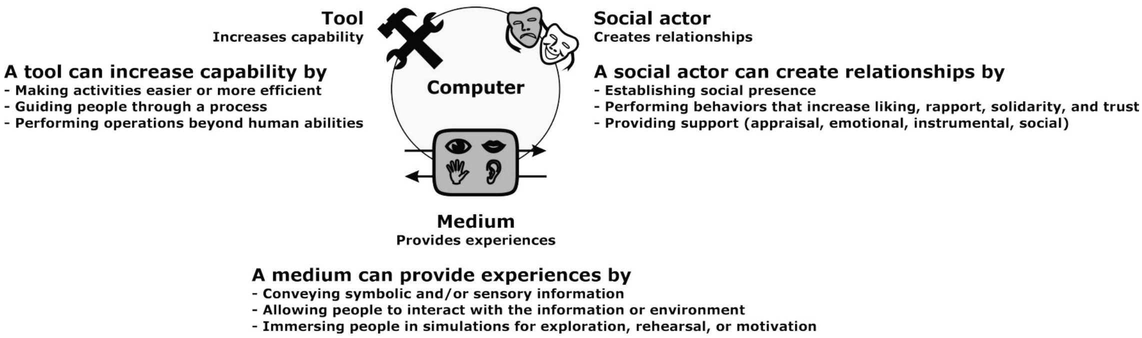

Figure 1. Different functional roles of computers and how they benefit the user. Adapted from Fogg (2003: 25, Figure 2.1).

- *Computers as social actors*. The computer creates relationships with users. It exhibits cues that trigger automatic and unconscious social responses in users, leading them to detect the (co-) presence of a social actor and to treat the computer as they would treat another human being (or some other living creature) (cf. Chapter 6.5). Further steps toward relationship building involve the computer showing verbal and non-verbal behaviors that promote liking, rapport, solidarity, and trust with the user. A third characteristic of computers as social actors is their ability to provide various kinds of support which people look for in relationships (e.g. advice, reassurance, assistance, social facilitation, etc.).

  The capacity of computing technology to lead users to perceive the presence of a social actor is an important prerequisite for the success of the principal actors whose design and roles are the concern of this thesis: *embodied interactive software agents*. These agents will be introduced in Chapter 3. When perceived as social actors, computers in general and agents in particular can influence users' attitudes, behaviors, motivation, etc. through their co-presence and interactions with users in social situations and through their ability to create relationships with them.

Today, computers fill more and more roles in society that traditionally have been filled by humans (Alvarez-Torres et al. 2001: 164f.), acting as bank tellers, receptionists, secretaries, market researchers, game characters, social companions, and teachers/tutors, to name but a few current 'professions' of computers. This thesis is concerned with the design of computer-based actors, or agents, for roles in *education* (cf. Chapter 16), in particular education delivered over the World Wide Web (cf. Chapter 2), although roles of agents in other domains will be discussed in Chapter 3.2.

1.2 User Interfaces

Not only the functions of computers are constantly evolving but also the ways in which humans interact with them. The interaction between human and machine occurs at the *user interface* (Shneiderman & Plaisant 2004), which consists of the facilities (hardware and software) by which people (the users) can exchange information and instructions with a particular computing device (the *system*). A user interface provides means of input (allowing users to control the system) and means of output (allowing the system to give users feedback by informing them about results, errors, and its internal state). Important types of user interfaces include the following:

- *Command-line interfaces (CLIs)* take parameterized commands in a formal language, which are provided as lines of text typed by the user on a keyboard, and execute them, presenting results as text output on the screen.
- *Graphical user interfaces (GUIs)* involve the direct manipulation of items on a graphical desktop using a pointing device. GUIs are essential components of state-of-the-art computer-based systems. After their central elements, they are also called *WIMP (Window, Icon, Menu, and Pointer) interfaces*.
- *Web-based user interfaces* are graphical user interfaces that accept input and provide output by generating web pages which are delivered over the Internet and viewed by users in a web browser program. The web pages use text, graphics, animations, audio, and video to present information. Furthermore, they may contain hyperlinks, form fields, buttons, and other input elements that allow the user to provide information and exercise control.
- *Tactile interfaces* use haptic devices like data gloves and joysticks for input and/or output. These devices enable the user to both provide information and to receive feedback from the computer in the form of a felt sensation on some part of the body (e.g. force feedback).
- *Touch screen interfaces* are graphical user interfaces that make use of a touch screen display as a combined input and output device. The user directly manipulates elements on a touch screen using his or her finger or a stylus. Touch screen interfaces may also feature advanced input options, such as handwriting recognition.
- *Conversational interfaces* are an emerging class of human-computer interfaces that allow their users to formulate requests in natural language, possibly combined with other input and output modes, in the context of an ongoing conversation between the user and the system.
- *Embodied conversational interfaces* add an animated (human-like) *character*, or *persona*, to the interaction, which represents the system or another user. This character can engage a human user in a real-time multimodal dialogue that emulates the experience of human face-to-face interaction, involving the coordinated use of verbal and non-verbal behaviors.

The general purpose of user interfaces is to hide the complexity of the underlying system from users by helping them to experience the system in a way which they perceive as effective, efficient, enjoyable, and satisfying. Ideally, an interface should facilitate *natural interaction* (Valli 2004), which is intuitive and does not rely on complex devices or extensive prior instruction of users. People interact naturally every day through physical actions enabled by the means provided by their bodies: speech, facial expressions, gaze, gesture, posture, and locomotion (cf. Chapter 10.1). Embodied conversational interfaces have the ability to understand these actions and can respond with the same kinds of behaviors, performed by embodied characters, thus allowing users to interact naturally with the system and with each other. The present research is concerned with embodied conversational interfaces and, to a lesser extent, with conversational interfaces. The focus is on the design of embodied, agent-based interfaces that are capable of multimodal (conversational) interaction with human learners.

1.3 Intelligent Software Agents

Human-machine interfaces have evolved considerably over the last fifty years. The cryptic command languages and outputs of the early days of computing have largely given way to the direct manipulation of on-screen objects in current graphical and web-based user interfaces. Another paradigm shift began in the early 1990s with the arrival of *intelligent software agents* in the world of computing. Discussed in detail in Chapter 3, intelligent software agents can be briefly characterized at this point as computer programs that have been told (or have learned) what a particular user needs, prefers, or wants, and have the ability to autonomously perform specific tasks which the user has delegated to them.

Human-computer interaction (HCI)[3] has always been (and still largely is) a master-slave (some would say user-tool) relationship, in which the user monitors and controls the operation of the computer, telling the machine at each step in the interaction what to do next. Unlike their users, current computers rarely take the initiative or act according to their own agenda in order to serve the user. In addition, for the most part, they do not possess the ability to adapt to an individual user; instead, people have to customize them. In other words, computers are *adaptable* rather than *adaptive* machines (Opperman 1994; Jameson 2003). Then again, state-of-the-art computer systems are not only sophisticated but also complex artifacts and have a constantly growing number of non-expert users. The intention and hope for intelligent software agents is that they can help to transform human-computer interaction from user-controlled command execution into a *cooperative process*, in which software agents adaptively, competently, and proactively assist users (and other agents) in various roles and domains (cf. Chapter 3.2).

The cooperation between humans and software agents is facilitated if agents are capable of making the interaction natural for the user, as outlined in the previous section. Hence, agents should be coupled with an embodied conversational interface that can handle (spoken or written) natural language, facial expressions, eye gaze, posture, and locomotion, both as inputs provided by users and as the agent's own outputs delivered by an animated, embodied character that portrays a rich individual and sociocultural identity and displays contextually and individually appropriate emotions. This thesis will discuss the design of both the front-end and the back-end components of embodied interactive agent systems for roles in education (referred to as *pedagogical agents*, cf. Chapter 4). The character-based front-end of these systems provides the body, face, and voice with which the pedagogical agent presents itself to and interacts with the learner, whereas the agent-based back-end implements the expertise and functionality that enables the agent to play its supporting role with respect to the learner and other participants (humans or other agents) in the learning experience.

1.4 E-Learning

The present work on embodied pedagogical software agents is embedded in the broader context of research and development involving the use of information and communication technologies (ICT) to facilitate and promote human learning. The broad spectrum of activities using ICT for educational purposes is commonly subsumed under the umbrella term *e-learning*. While there is no generally accepted definition of e-learning, in this work, it is taken to involve the creation, delivery, promotion, and facilitation of individual and collaborative learning processes and experiences situated in a learning environment built on computer and network technologies, in particular the World Wide Web ([INT 2]). These environments integrate communicative, interactive, and multimedia components for content management, content delivery, learning

---

[3] Human-computer interaction (HCI) is concerned with the design, implementation, and use of computer interfaces (Jacko & Sears 2003). The goal of HCI is to create effective and enjoyable systems that help people to accomplish their individual or collaborative tasks.

management, communication, collaboration, and student assessment. A more detailed account of e-learning will be given in the next chapter.

The specific web-based learning environment providing the context for the present work is the Virtual Linguistics Campus (VLC) ([INT 3]; Handke & Franke 2006), the world's leading e-learning platform for linguistics, which offers a wide variety of web-based courses in the fields of theoretical linguistics, applied linguistics, language technologies, teacher training, and linguistic fieldwork. The linguistic fieldwork classes of the VLC serve as the environment for three pedagogical agent roles, which will be discussed in Chapter 16.

1.5 Overview of the Thesis

This thesis consists of seventeen chapters, which can be grouped into chapters on basic concepts including e-learning, intelligent software agents, pedagogical agents, and interactivity (Chapter 1 to Chapter 5); chapters on design, design principles for embodied interactive software agents, and agent design failures (Chapter 6 to Chapter 8); chapters on the design of different components of embodied interactive pedagogical agents (Chapter 9 to Chapter 13); chapters on agent architectures and technical platforms as well as on the evaluation of embodied interactive software agents (Chapter 14 to Chapter 15); a chapter which discusses three different pedagogical agent roles for the linguistic fieldwork classes on the Virtual Linguistics Campus (Chapter 16); and a final chapter summarizing the thesis (Chapter 17). The contents of the chapters following this introduction are summarized below:

- *Chapter 2* defines and delimits the scope of the term 'e-learning;' compares three major theoretical perspectives on human learning (behaviorism, cognitivism, and constructivism) and discusses the implications of each perspective for the design of instruction; reviews a number of e-learning technologies relevant to the present research; and provides an overview of the structure, components, and features of the Virtual Linguistics Campus.
- *Chapter 3* is concerned with intelligent software agents. It defines the concept and shows what distinguishes software agents from other types of computer programs; discusses the properties of software agents in detail; and describes different types of intelligent software agents. A separate section gives an overview of the qualities and components of *embodied interactive software agents*, i.e. intelligent software agents with a digital embodiment that enables them to act on and react to other entities (users, objects, and other agents) through bodily actions, including verbal and non-verbal communicative behaviors.
- *Chapter 4* introduces pedagogical agents as embodied interactive software agents designed to provide various kinds of assistance to human learners in the context of interactive learning environments. Furthermore, it describes the capabilities which enable pedagogical agents to facilitate learning and the requirements that pedagogical agents with these capabilities have to meet.
- *Chapter 5* discusses the notion of *interactivity*, which is widely regarded as a positive quality that artifacts, exchanges, and experiences should possess, although its precise nature and effects are not known (Bucy 2004b: 373f.). This is an unsatisfying situation for the present research, given that embodied *interactive* software agents by their very definition are capable of engaging in mutual give and take with other entities. To develop a better understanding of interactivity for agent design, this chapter examines the notion in some detail, revealing that it is more complex than it appears on the surface, but also contributing a number of useful

insights for the design of interactive agents in general and for roles in e-learning in particular.

- *Chapter 6* provides the necessary background on the field and process of design for the discussion of agent design in later chapters. It begins with a discussion of the nature and characteristics of the user as the individual who makes use of designed artifacts. The context which influences the use and usability of artifacts, including software agents, is discussed next. The remainder of the chapter consists of separate sections covering cognitive, affective, and social aspects of design. Fundamental concepts, including the usability of artifacts, the interplay between cognition and affect and its influence on users' experience of artifacts, and the Media Equation, are discussed in these sections.
- *Chapter 7* argues that the design of embodied interactive software agents for education and other domains should be based on solid principles rather than intuition and ad-hoc decisions. Sixteen design principles are proposed in this chapter: Added Value, Perceptible Qualities, Balanced Design, Coherence, Consistency, Completeness, Comprehensibility, Individuality, Variability, Communicative Ability, Modularity, Teamwork, Participatory Design, Role Awareness, Cultural Awareness, and Relationship Building.
- *Chapter 8* analyzes two less than successful earlier designs of embodied interactive software agents, Microsoft Bob and the Office Assistant, from the perspective of the design principles discussed in Chapter 7, showing that their failure results from a lack of consideration for several of these principles and deriving a number of lessons for the design of embodied interactive agents. The chapter also provides the transition to the following chapters covering the different components of agent design (Chapter 9 to Chapter 13).
- *Chapter 9* is concerned with the first (and most immediately visible) aspect of the design of an embodied interactive software agent: the body with which the agent presents itself to the user. It discusses the pros and cons of humanoid embodiments, several guidelines for designing an agent's appearance, the need for individual agent designs based on a character profile, the ability of the embodiment to move, the appropriate degree of realism for the agent's representation, applications of complete and partial embodiments, and the issue of whether an agent should be able to change its embodiment.
- *Chapter 10* discusses the design of different kinds of behaviors displayed by an embodied interactive agent during its interactions with the user. Categories of behaviors involved in communication are the topic of the first section, including voice, gesture, facial expression, gaze, posture, and locomotion. The second section covers behaviors performed by an agent to enhance the believability of its persona. In the third section, behaviors that help to build agent-user relationships are discussed. Finally, the fourth section addresses the issue of the appropriate degree of presence of embodied agents on the screen (high vs. low profile).
- *Chapter 11* deals with the conversational abilities of embodied interactive software agents. The discussion begins with reviews of the state of the art in the technologies involved in giving (embodied) conversational agents the ability to understand and produce (spoken) natural language and to process multiple combined input modes. The focus of the remainder of the chapter is then on the design aspects of different types of conversational agents, including chatbots, dialogue agents, and embodied conversational agents.
- *Chapter 12* is concerned with the different components of the expertise of a pedagogical software agent. Following the classic subdivision made in intelligent tutoring systems, the chapter first discusses the agent's domain model, including techniques for the acquisition, representation, and use of knowledge; then the student model, covering user modeling and student modeling in general as well as comparing different approaches to student modeling;

and finally the instructor model, describing various instructional interactions that embodied pedagogical agents can implement to assist learners.

- *Chapter 13* shifts the focus of attention from the cognitive, 'rational' aspects of an agent to its "inner life." The major part of the chapter is concerned with the notion of 'emotion' in relation to embodied interactive software agents. The discussion covers the dual (cognitive and physical) nature of emotion, its functions in cognition and social interaction, communication-driven and simulation-based approaches to modeling emotion in machines, and ways for agents to interpret the emotional state of the user and to express their own (simulated) emotions. The two remaining sections of the chapter deal with the modeling, projection, and adaptation of the agent's personality and discuss the need to add depth to the agent's personality by enriching it with a detailed back-story.
- *Chapter 14* is concerned with technical aspects of agent design. It compares different basic architectures for software agents, including deliberative, reactive, hybrid, blackboard, and multi-agent architectures. Furthermore, it discusses four approaches to generate believable behaviors of embodied agents: behavior sequencing, layered generation, state machine compilation, and multi-character systems. Finally, different technical platforms for software agents are compared, including desktop computers, the World Wide Web, virtual reality, and mobile devices.
- *Chapter 15* discusses aspects of the evaluation of embodied interactive software agents. It begins by identifying motivations for doing evaluation research on and with these agents. One important motivation is the possibility to derive design guidelines from the results of evaluation research. Factors that can influence the evaluation of agents are discussed next. They include design parameters of the agent, characteristics of the user, and contextual factors. The following sections review different types of evaluation research (formative and summative, macro and micro evaluation), discuss basic aspects of evaluation research design (the research model, the research question, and constructs), and compare research strategies, data collection methods, and qualitative and quantitative data analysis, respectively. The chapter concludes with a section providing a set of guidelines for the evaluation of embodied interactive software agents.
- *Chapter 16* describes three different roles for embodied interactive pedagogical software agents in the context provided by the linguistic fieldwork classes of the Virtual Linguistics Campus. The three roles include the *virtual native speaker*, which serves as an interactive informant for the learner's field research, the *coach*, which offers guidance and motivational support during the learner's fieldwork activities, and the *peer*, which acts as a co-learner exploring the language together with the learner. The design and evaluation of the role of the virtual native speaker are discussed in detail.
- *Chapter 17* summarizes the main results of the thesis and provides an outlook on future work on the design, implementation, and evaluation of the pedagogical agent roles described in Chapter 16.

1.6 Contribution of the Thesis

The goal of the present work is both theoretical and practical in nature. One the one hand, its intention is to contribute to a scientific understanding of issues, methods, theories and technologies involved in the design, implementation, and evaluation of embodied interactive software agents for different roles in education. On the other hand, it aims to provide guidelines for practitioners, which help them to build agents for education and other application domains that effectively play the (instructional) role for which they have been designed and are perceived by users as believable and competent artificial individuals. The main contributions of the thesis are the following:

- A comprehensive, multidisciplinary, and user-oriented view of the design of embodied interactive software agents, which brings together inputs from a variety of both academic and non-academic sources, including the arts, computer science, linguistics, psychology, pedagogy, robotics, graphics and animation, the social sciences, computer game research and development, the science-fiction literature, and (animated) movies, all of which have made (intentionally or unintentionally) some kind of contribution to the understanding of what is involved in being human. The insights into the design of embodied interactive software agents compiled in this thesis have applications not only in the development of pedagogical agents but also in the design of embodied interactive agents for other application domains.
- Sixteen principles and numerous specific guidelines for the design of embodied interactive software agents and their components.
- A critical review of theories, concepts and approaches from various areas and disciplines that are relevant to the design of embodied interactive pedagogical agents, in preparation for future development work, which is outlined in Chapter 17.
- The discussion of a concrete application scenario on the Virtual Linguistics Campus for embodied pedagogical software agents in three different roles (virtual native speaker, coach, and peer), including a detailed description of the design of one of these roles (the virtual native speaker).

2 E-Learning

Education needs to prepare individuals for living and working in a world that depends on information and knowledge like humans depend on oxygen, and, despite the essential support provided by technology, still builds on the ability of the individual to acquire, understand, manipulate, create, update, refresh, publish, and preserve information and knowledge. Learning has become a necessary lifelong activity because knowledge and skills have to be refreshed, updated, and extended at an ever increasing rate until post-retirement. Furthermore, the amount of information available on the Internet and elsewhere grows exponentially and has to be handled by individuals, the society, and the education system without becoming overwhelmed by it. In addition, people increasingly need to be equipped "just in time" or "on demand" with the knowledge and skills necessary to handle a given problem or situation. As a result, education and training are nowadays usually expected to be flexible with respect to the place and time of learning. Finally, state-of-the-art *information and communication technologies (ICT)*, such as computers, cell phones, and portable media players, have transformed the ways in which people work, play, and communicate. Especially for the younger people from the so-called *Net Generation*[4] (Tapscott 1998; Oblinger & Oblinger 2005) in first-world societies, who were born in the 1980s or later, have grown up with ICT, and are now starting out in their academic or vocational career, these are familiar artifacts, which they naturally use every day for communication, entertainment, and, recently, also learning. It is both a responsibility of and an opportunity for the education system to integrate ICT into learning scenarios in order to help people to learn more effectively and to develop the competence they need to make effective use of ICT as early as possible in life.

The use of information and communication technologies to deliver, manage, and support learning experiences, usually in a networked setting, is often subsumed under the general label of *electronic learning* or *e-learning*. In recent years, e-learning has been proposed as an answer to the challenges that education faces in the information age, as outlined above. Comprehensive reviews of e-learning can be found elsewhere (Kearsley 2000; Ko & Rossen 2001; Rosenberg 2001; Clark & Mayer 2003; Mason & Rennie 2006; Schulmeister 2006). Instead, the following discussion introduces a number of basic concepts, technologies, and scenarios related to e-learning in order to provide the necessary background for the discussion in later chapters. Section 2.1 presents an elaborate definition of e-learning and delimits the scope of the term for the present work. Since e-learning requires a solid foundation in principles of human learning, Section 2.2 discusses learning and instructional design in the context of three major theoretical perspectives: behaviorism, cognitivism, and constructivism. Section 2.3 reviews technologies for e-learning. Section 2.4 describes the Virtual Linguistics Campus, the e-learning platform that provides the context for the present research. Finally, Section 2.5 summarizes the chapter.

2.1 Definition and Scope

> *The **e** in e-learning stands for **e**lephant. Or at least it might. For like the blind men encountering an elephant in the old story, e-learning means different things to different people. One feels the*

---

[4] It should be noted that the Net Generation is a quite controversial notion. In the German e-learning community, Rolf Schulmeister is one of the most renowned and outspoken critics of this idea (Schulmeister 2008).

tail and exclaims, "e-learning is like a T-3 line to the Internet backbone." Another is slapped in the face by the trunk and says, "e-learning is about interactivity and feedback." Yet another touches the elephant's side and says, "Whatever e-learning is, it's very, very big." (Jay Cross, [INT 4], his emphasis).

The term 'e-learning' covers a wide range of different uses of information and communication technologies for educational purposes. Its meaning has changed since it first appeared in the late 1990s (coined by Jay Cross ([INT 5])), and it has different connotations in different communities and contexts. Definitions of e-learning abound in the literature and on the World Wide Web (Romiszowski 2004). Some of these definitions emphasize content, others communication, and still others technology (Mason & Rennie 2006: xiv). While a number of views classify virtually any form of technology-supported learning (including the use of audio and video tapes, CD-ROMs, and satellite broadcasts to deliver educational content to learners) as 'e-learning', others reserve the term for educational scenarios that use the Internet as a platform (p. xiv), like, for example, Rosenberg:

E-learning refers to the use of Internet technologies to deliver a broad array of solutions that enhance knowledge and performance. It is based on three fundamental criteria:

1. *E-learning is networked, which makes it capable of instant updating, storage/retrieval, distribution and sharing of instruction or information. (...)*
2. *It is delivered to the end-user via a computer using standard Internet technology. (...)*
3. *It focuses on the broadest view of learning – learning solutions that go beyond the traditional paradigms of training. (Rosenberg 2001: 28f.).*

Not only does e-learning mean different things to different people, as pointed out in the introductory quotation, but there is also a plethora of other terms which are related to e-learning and often confused with it, although they have (more or less) different meanings (Mason & Rennie 2006: xvf.). Examples include *web-based learning*, *distance education*, *e-education*, *distributed learning*, *flexible learning*, *open learning*, *virtual classroom*, and *computer-based/-aided/-assisted/-mediated education/instruction/learning/training*, among many others. Guri-Rosenblit correctly observed that:

Often discourse on new technologies suffers from "The Tower of Babel Syndrome" – a confusing language and misleading conclusions, emanating from the fact that people refer to totally different functions and roles while using the same generic terms (...). It is not just the meanings attached to specific terms that are unclear. The language used in the relevant literature to depict the nature of study environments shaped by the new technologies is blurred and confusing. (Guri-Rosenblit 2005: 6).

To be as clear as possible about the meaning of the term 'e-learning' as it will be used in the present work, a definition is proposed below that is based on the author's first-hand experience of creating and teaching several e-learning courses, both with and without face-to-face interaction, in the learning environment of the Virtual Linguistics Campus, and of programming a substantial part of the underlying technical framework of this environment:

E-learning is the use of Internet technologies to create, deliver, and support individual and collaborative learning experiences and processes anytime and anywhere, with or without face-to-face components and involving human (or non-human) staff to different degrees and in different roles (teacher, tutor, coach, guide, mentor, etc.). It integrates administrative, communicative, interactive, and multimedia elements into a computer-based, networked learning environment that is adaptive, dynamic, and individualized with respect to the content and support it provides. The design of the educational content, the learning support, and the environment as a whole is based on principles of instructional design and guided by formative and summative evaluations.

This definition is intentionally elaborate and comprehensive in order to cover as broad a spectrum of e-learning activities as possible. One focus of the definition is on the technical side of e-learning, in particular the components and properties of e-learning environments that use the Internet, especially the World Wide Web, as a platform. These environments have tasks which go beyond the delivery of educational content and include the provision of dynamic and tailored support for both the individual learner and groups of learners. The definition covers e-learning scenarios that are, in principle, independent of place and/or time but may combine online learning at the computer via the Internet with face-to-face sessions in the real world. Human instructors, but increasingly also non-human teaching and support staff, i.e. autonomous pedagogical software agents (cf. Chapter 4) with characteristics including embodiment, behavior, communicative abilities, expertise, emotion, and personality (cf. Chapter 9 to Chapter 13), may be involved in these scenarios to different extents and in different roles. The process of developing the educational content, the mechanisms for supporting learning, and the overall environment has a solid foundation in principles for systematically planning, designing, and evaluating educational content and environments, and is constantly informed by repeated evaluations of increasingly sophisticated versions of the learning environment throughout the development process (cf. Chapter 15).

Figure 2 shows the relationship between e-learning and a number of associated terms that will be discussed later in this chapter, as it is perceived by the author (other views may differ). In this diagram, e-learning both (partially) subsumes and is partially included in other terms, which are described briefly below:

- *Face-to-face* refers to education that takes place when learners and instructors have direct, unmediated contact, i.e. when they interact with each other in the same physical space and at the same time. Education delivered in the classroom or in one-to-one tutoring sessions are classic examples of face-to-face scenarios. The emphasis of face-to-face instruction is on the interaction between learners and instructors; use of technology and (online) educational material is supplemental to human teaching.
- *Online learning* (also called *web-based learning*) (Drave 2000; Kearsley 2000; Ko & Rossen 2001) does not involve any face-to-face interaction between the participants but relies solely on information and communication technologies to deliver content, provide assessment, and offer support to learners, and to provide the tools for communication between and among learners and instructors (Mason & Rennie 2006: 126). Current online learning scenarios commonly use the World Wide Web as their technical platform. A typical web-based course has the following features (p. 126, cf. Section 2.4):

    - Instructional material, preferably in an interactive multimedia format;
    - Course description, dates and deadlines, course requirements;
    - Facilities supporting formative and summative assessment of learners;
    - Learner management tools (records, statistics, tracking, and modeling);
    - Communication tools (discussion forums, e-mail, and chat);
    - Links to further internal and external resources (databases, libraries, web sites, etc.).

    In addition, more recent online courses integrate tools such as weblogs and wikis to involve learners as co-authors of content (cf. Section 2.3.9).
- *Blended (or hybrid) learning* (Bersin 2004; Bonk & Graham 2006) integrates elements of face-to-face and online learning scenarios, combining the flexibility ("anytime, anywhere") of online delivery with the social environment of the classroom (Mason & Rennie 2006: 12f.). For example, blended learning on the Virtual Linguistics Campus (cf. Section 2.4.2.3)

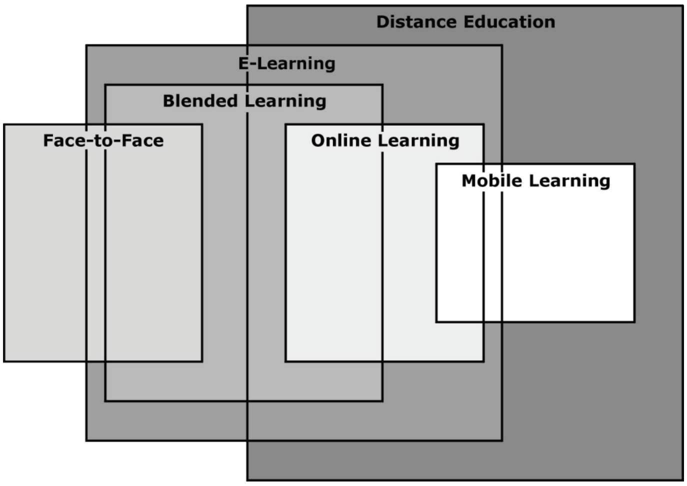

Figure 2. The relationship between e-learning and various associated terms. Based on Mason and Rennie (2006: xvii, Figure 1).

involves learners studying online educational content before participating in in-class sessions with the instructor that focus on discussion and practice rather than presentation of content. In general, blended learning is now regarded by many as superior to both face-to-face and online scenarios (e.g. Dean et al. 2001; Singh 2003; Dziuban et al. 2004; Frankle 2005).

- *Mobile learning (or m-learning)* (Keegan 2002; Alexander 2004; Naismith et al. 2004; Sharples 2007) can be broadly defined as "any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of learning opportunities offered by mobile technologies" (O'Malley et al. 2003: 6). M-learning is a recent buzzword (Mason & Rennie 2006: xvii) referring to ways of delivering educational content to the "mobile generation" in a format that is appropriate for cell phones, portable media players, and other mobile devices. Mobile learning shares some aspects with online learning, but educational content is typically downloaded to the mobile device (e.g. as a *podcast*, cf. Section 2.3.9) rather than studied online.
- *Distance education* (Moore & Kearsley 1996; Moore & Anderson 2003; Neal & Miller 2005) refers to scenarios in which learners are separated from their instructors and peers in terms of location and often also time (Neal & Miller 2005: 454). Starting with the first correspondence courses in the 1800s, various technologies have been used to deliver instruction to remote learners, including radio, television, computer-based courseware on CD-ROMs, audio- and video conferencing, and, recently, web-based courses and mobile devices (p. 455). While distance education implies that teaching and learning processes are separated by place and/or time, some face-to-face interaction may be involved (e.g. kick-off meetings, on-site workshops, and in-person exams). Distance education includes both online learning and mobile learning but is broader than both.

2.2 Learning and Instruction

Learning (Pear 2001) is the complex cognitive process by which people[5] acquire and develop knowledge and skills from sources including experience, practice, study, and instruction (OED

[5] See Chapter 12.1.1.2 for a discussion of *machine* learning.

1998: 1048; [INT 6]). It is an intra-individual process that is influenced by a variety of factors, including the individual's mental and physical abilities, previous knowledge and experiences, current affective state, personality, and his or her physical and social environment. Learning induces relatively permanent changes in the individual's actual behavior or his or her capacity for behavior (Schunk 1991: 2).

Instruction is the deliberate process of initiating, promoting, facilitating, guiding, and assessing the learning processes of individuals and groups, which is working toward achieving specific learning goals identified for them (Smith & Ragan 2005: 4). In contrast to learning, instruction takes place *between* individuals. It targets learners individually or as a group from the outside in order to affect changes within each learner. Instruction is often used synonymously with the terms *education*, *training*, and *teaching* (p. 5). However, following Smith and Ragan (pp. 4ff.), in the present work, these terms are defined as related to, but also distinct from, instruction (cf. Figure 3). *Education* is the broadest category, which includes all experiences in which learning occurs, whether they are planned or unplanned, formal or informal, intentional or incidental. Instruction is completely included in education since all instructional activities are performed to promote learning. The reverse is not true, though, because people learn from many experiences that are not explicitly designed to be instrumental in the achievement of specific learning goals. While instruction is often reduced to strategies for the transfer of knowledge and skills from a knowledgeable source (e.g. the teacher, a textbook, or an e-learning course) to learners, the definition of instruction above also applies to student-centered learning experiences, provided that they are directed toward specific learning goals. *Training* is instruction intended to help learners to acquire specific skills for which they have an (almost) immediate application. Often, training equips individuals with specific competencies which they need for their work in business, military, or governmental environments. However, training is only a subset of instruction because much of instruction does not prepare learners for accomplishing particular work-related tasks but has a more general scope. Furthermore, the knowledge and skills acquired are not for immediate application but may serve the individual on various occasions in his or her later life. Finally, the term *teaching* labels instruction that is provided by a live human teacher rather than by some kind of medium (e.g. a textbook, CD-ROM-based courseware, or an e-learning environment). It is important to note that in contrast to traditional views that equate teaching and instruction and consider the presence of a human teacher to be mandatory, the diagram in Figure 3 shows that instruction is larger than teaching and includes experiences for which a human teacher is not necessary and can therefore be replaced with computer-based support, such as pedagogical software agents. On the other hand, teaching is not always aiming at pre-defined instructional goals. Instead, learning goals may emerge during activities designed and implemented by the teacher, often developed by learners as they are engaged in an activity, producing possibly unanticipated learning outcomes.

Learning is a complex phenomenon that has attracted the attention of scholars for a long time, generating a vast body of theoretical insights from different perspectives (psychology, pedagogy, philosophy, linguistics, artificial intelligence, etc.). However, the results of basic learning research are usually not directly applicable to instructional practice but have to be translated into instructional strategies, techniques, and materials as well as guidelines for selecting the appropriate approach for a particular learner or group of learners in a specific educational context (Ertmer & Newby 1993: 50). *Instructional design (ID)* (Gagné et al. 2004; Smith & Ragan 2005) serves a bridging function between learning research and instructional practice (Ertmer & Newby 1993: 50) by "translating principles of learning and instruction into plans for instructional materials, activities, information resources, and evaluation" (Smith & Ragan 2005: 4). ID is a systematic, iterative, and reflective process (p. 4) that consists of

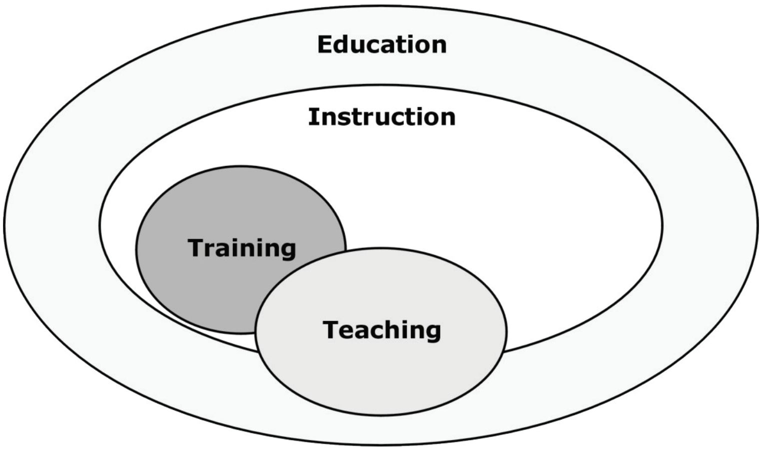

Figure 3. Instruction and related terms (Smith & Ragan 2005: 5, Figure 1.1).

identifying learning needs and goals for the target audience; designing materials, activities, resources, systems, etc. to address these needs and goals; evaluating the designs created; and improving them based on the results of these evaluations ([INT 7]). Products of instructional design not only have to be functional with respect to their educational purpose but also need to appeal to the learner (Smith & Ragan 2005: 4), a requirement that was often forgotten in the design of early computer software for learning. Affective, cognitive, and social aspects of design are discussed in Chapter 6.

Instructional design can draw important insights from an understanding of human learning processes. There are many different views and theories about how learning can be effectively achieved (Ertmer & Newby 1993: 51). *Learning theories* provide theoretical frameworks or models for the description and explanation of learning processes in humans and animals ([INT 8]). Learning theories differ from each other with respect to the learning mechanisms postulated; the factors that are said to influence learning; the role that memory is assumed to play in the learning process; the ways in which learners supposedly transfer knowledge and skills to new problems and situations; and the types of learning for which the theory offers the best explanations (p. 53). Three major theoretical perspectives on human learning and their implications for instructional design with and without technology are reviewed in the following: behaviorism, cognitivism, and constructivism.

2.2.1 Behaviorism

Popular in the 1950s and 1960s, behaviorism contends that learning of animals and people can be explained by only focusing on their objectively observable behaviors while the individual's internal cognitive structures and processes are not considered ([INT 9]) (i.e. the mind is viewed as a "black box"). At the heart of behaviorism is the *stimulus-response model*, which explains the observable behavior of an organism only in terms of the organism's reaction (response) to some previous stimulus. The process in which an organism learns to associate a certain stimulus with a certain response is referred to as *conditioning*. Hence, behaviorism views learning as an observable change of behavior which emerges in reaction to environmental stimuli and as a result of conditioning processes (Arnold 2005: 2). In the *classical* approach to conditioning (Pavlov 1927), a neutral stimulus (e.g. a light or a sound) is repeatedly paired with another stimulus (e.g. food) that evokes a reflex reaction (e.g. saliva production), until eventually the two stimuli become associated and produce the same response ([INT 10]). In contrast, *operant* *conditioning* (Skinner 1938) is the process of modifying the behavior of an organism through repeated *reinforcement* of desired responses and *punishment* of undesired ones. Reinforcement and punishment are either 'positive' (i.e. a pleasant or unpleasant stimulus is added following a response) or 'negative' (i.e. an unpleasant or pleasant stimulus is removed after a response), with positive reinforcement being more effective than punishment ([INT 11]; Arnold 2005: 3). By selectively reinforcing (or punishing) behaviors, the 'instructor' can completely control the learning process from the outside (p. 2). Behaviorists regard knowledge as given and absolute (Dara-Abrams 2002); the role of memory in learning is typically not addressed (Ertmer & Newby 1993: 55).

Instruction based on behaviorist principles emphasizes the attainment of observable and measurable learning outcomes, pre-assessment of learners to determine the starting point for instruction, simplification and standardization in instruction, decomposition of knowledge into basic building blocks, sequencing of instruction (cf. Chapter 12.3.11), mastery of the current item of instruction as a prerequisite for moving on to the next, reinforcement of desired responses through rewards and feedback, and use of cues, practice situations, and environmental conditions that strengthen stimulus-response associations (Ertmer & Newby 1993: 56f. + 59). The goal of instruction is to transfer knowledge to learners as efficiently and effectively as possible (Bednar et al. 1992).

Based on the ideas of behaviorism, Skinner developed the model of *programmed instruction* (Skinner 1968; [INT 12]), which, in turn, provided the theoretical foundation for the design of the first generation of programs for *computer-based training (CBT)* (Arnold 2005: 6). In these programs, the content of instruction is divided into small units, which are presented in a logical sequence ([INT 12]). Each unit is followed by questions or exercises for learners to determine their understanding of the material. Correct responses generate some positive feedback from the program and 'qualify' the learner for the next unit, whereas incorrect solutions have the opposite effect: the system states that the solution is incorrect, presents the correct answer and/or further information, and makes the learner repeat the unit or the exercise. Hence, the program completely controls the learning process, leaving the learner little or no freedom to act on his or her own. *Drill-and-practice programs* are classic examples of this approach (cf. Section 2.3.1) (Arnold 2005: 6).

For some time during the 1950s and 1960s, behaviorism was the most influential approach in behavior and education research, and its proponents claimed that it provided a universal explanation for learning in animals and people (Arnold 2005: 3). However, behaviorism soon came under attack for neglecting the interior of the "black box" (the mind), i.e. of the cognitive structures and processes involved in perception, memory, problem solving, etc., and for reducing learning to *observable* changes of behavior (p. 3).

Educational software based on behaviorist principles is appropriate for reinforcing simple skills and recollection (Bostock 1995), which are involved in vocabulary drills, basic calculation exercises, and spelling tasks, for example. However, these programs are not suitable for more complex content or skills, and they do not help students to establish connections between different pieces of knowledge, teach them how to apply newly acquired skills and knowledge in realistic scenarios, or encourage them to engage in critical reflection on their own knowledge (Arnold 2005: 7).

2.2.2 Cognitivism

Toward the end of the 1950s, the growing discontent with behaviorism marked the beginning of the *cognitive revolution* in psychology (Broadbent 1958; Chomsky 1959; Neisser 1967), which brought about a paradigm shift from the behaviorist view of the mind as a "black box" to a consideration of the internal cognitive structures and processes of the individual. Phenomena such as perception, memory, thinking, problem solving, and language processing became the focus of cognitive theories about how complex mental concepts and structures are acquired (Arnold 2005: 3). Cognitivists agreed with behaviorists that learning should be studied objectively by conducting empirical research. But unlike behaviorists, they tried to explain observable behavior in terms of its underlying internal cognitive structures and processes ([INT 13]). Ertmer and Newby summarized the key points of cognitive research on learning as follows:

> *Cognitive theories stress the acquisition of knowledge and internal mental structures; they focus on the conceptualization of students' learning processes and address the issues of how information is received, organized, stored, and retrieved by the mind. (...) Learning is concerned not so much with what learners **do** but with **what** they know and **how** they come to acquire it (Ertmer & Newby 1993: 58, their emphasis).*

In the cognitive tradition, learning is viewed as mental *information processing* that works in a similar way to the input-processing-storage-output architecture of computers (Driscoll 2000: 75; Siemens 2006): some input from the environment (e.g. explanations, demonstrations, worked examples, corrective feedback, etc.) enters short-term memory, where it is processed. The processed information is stored in long-term memory for later retrieval and/or results in some sort of (changed) behavior (output). The learner actively develops and repeatedly modifies cognitive structures by applying cognitive processes to incoming information; this is called *active learning* (Mayer 2005c: 36). These processes either change learners' existing cognitive structures to adapt them to new experiences (*accommodation*), which can also involve the creation of new structures, or incorporate new experiences into existing cognitive structures (*assimilation*) (Piaget 1952). The goal is to overcome the individual's state of *perturbation* (caused by unmet expectations about the environment) and to create a new equilibrium (Schulmeister 1997: 71).

Cognitive research distinguishes between different kinds of knowledge, including *factual knowledge* (of basic terminology and specific facts), *conceptual knowledge* (of categories, principles, and models), *procedural knowledge* (of procedures, techniques, and methods as well as criteria for their application), and *metacognitive knowledge* (of one's own knowledge and capabilities) (Mayer 2002: 625f.). Knowledge is assumed to exist independently of the individual; in other words, cognitivists (like behaviorists) see knowledge as given and absolute (Dara-Abrams 2002; Arnold 2005: 4). There is a single 'correct' *mental model* (knowledge structure) capturing the central concepts of the subject matter and the relationships between them, which is formed by the learner through active cognitive processing (Gentner & Stevens 1983; Dalgarno 1996; Mayer 2005c: 36). Examples of basic knowledge structures include process, comparison, generalization, enumeration, and classification structures (Mayer 2005c: 36). The task of instruction is to support learners' construction of mental models, by helping them to process new information and integrate it with what they know in a way that facilitates later retrieval of relevant knowledge for transfer to a new situation or task (Clark & Mayer 2003: 35–42). How this can be accomplished in the instructional design of multimedia learning experiences is discussed in the following within the framework of Mayer's *cognitive theory of multimedia learning (CTML)* (Mayer 2005c).

The goal of this theory is to explain human learning by referring to "how mental processes transform information received by the eyes and ears into knowledge and skills in human memory" (Clark & Mayer 2003: 35). *Multimedia learning* involves constructing mental models from instructional materials that combine *words* (spoken or printed) and *pictures* (static or moving images), i.e. auditory and visual information (Mayer 2005b: 2). Three assumptions about human cognitive processing are central to the CTML (Mayer 2005c: 33–37):

- *Dual channels*. Humans are equipped with two separate channels for processing auditory and visual information, respectively: an *auditory/verbal channel* and a *visual/pictorial channel*.
- *Limited capacity*. There are limits to the amount of information that can be held and processed in each channel at any one time.
- *Active processing*. Learning involves *selecting* relevant incoming information on the auditory and the visual channels, *organizing* the selected information from each channel into a coherent mental representation for that channel, and finally *integrating* the mental representations of the auditory and the visual information with each other and with existing knowledge (cf. Figure 4).

Three memory stores play a role in cognitive learning processes: visual and auditory information is perceived through the eyes and ears, briefly held in *sensory memory*, processed in *working memory*, and finally stored in *long-term memory* (Clark & Mayer 2003: 35). The three stores are briefly described below (Mayer 2005c: 37f.):

- *Sensory memory* stores perceived auditory and visual information separately as exact auditory and visual images for a very short time before they enter working memory.
- *Working memory* is the cognitive structure in which conscious processing of information (including learning) takes place (Sweller 2005a: 29). It provides a capacity for storing and manipulating auditory and visual information; however, this capacity is severely limited with respect to how much information can be held and for how long (p. 29). "Seven plus or minus two chunks" is an often-heard estimate of the number of elements that can be stored in working memory at any one time (Miller 1956). The number of elements that can be simultaneously processed (i.e. combined, contrasted, or manipulated) is even smaller (2–4) (Sweller 2005a: 21). New information is retained only for a few seconds if it is not rehearsed (Peterson & Peterson 1959).
- *Long-term memory* stores knowledge in large quantities and for a long time (Mayer 2005c: 47). For learning to occur, new information in working memory has to be integrated with prior knowledge from long-term memory and *encoded* into long-term memory after active processing (*rehearsal*) in working memory. The encoded information has to be *retrieved* from long-term memory back into working memory when needed in later cognitive processing (Clark & Mayer 2003: 36).

Hence, the cognitive theory of multimedia learning regards learners as information processors with certain inherent capabilities and limitations that have to be taken into account in instructional design. Sweller observed that "[g]ood instructional design is driven by our knowledge of human cognitive structures and the manner in which those structures are organized into a cognitive architecture" (Sweller 2005a: 19).

Instructional design in the cognitive tradition aims to help learners to transform the incoming auditory and visual information from the learning experience, using the storage and processing capacities provided by their sensory and working memories, in order for the processed

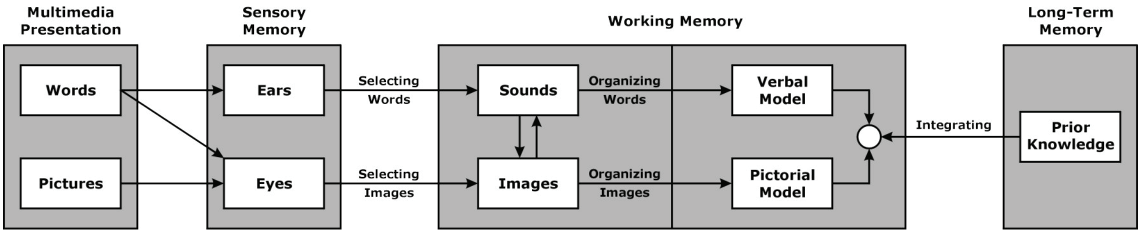

Figure 4. Mayer's cognitive theory of multimedia learning (Mayer 2005c: 37, Figure 3.2).

information to be eventually integrated into the knowledge structures stored in their long-term memory (Clark & Mayer 2003: 36). This transformation involves five cognitive processes (Clark & Mayer 2003: 36–42; Mayer 2005c: 38–41):

1. *Selecting relevant auditory and visual information for processing*. Learners can identify pertinent information sources more easily if their attention is guided by appropriate instructional methods and design techniques. For example, e-learning units may provide a list of learning objectives in their introductory section and use arrows, colors, font size and weight, and other visual cues to highlight important aspects of the information presented.
2. *Managing limited working memory capacity*. Working memory can only store a very limited amount of information at any given time and is therefore a scarce resource in the learning process. The burden on working memory consists of both information to be held and information to be processed and is commonly referred to as *cognitive load* (Sweller 2005a). The overburdening of working memory leads to *cognitive overload*. By reducing cognitive load in learning experiences, valuable capacity in working memory becomes available for the active processing of important information, which is the prerequisite for learning, and cognitive overload can be avoided. Ways to reduce cognitive load include off-loading (reassigning) information from the visual to the verbal channel (e.g. presenting words as narration rather than text), eliminating (weeding) interesting but extraneous (irrelevant) material, and synchronizing the presentation of auditory and visual material spatially and temporally (Mayer & Moreno 2003).
3. *Rehearsing and integrating information*. The selected auditory and visual items are organized into different coherent knowledge structures by building structural relations among the auditory and among the visual elements, respectively. The next step is to integrate the resulting auditory and visual representations in working memory, first with each other and then with the learner's prior knowledge retrieved from long-term memory (Mayer 2005c: 40f.), by creating connections among the new elements and between new and existing elements. This active processing (rehearsal) takes place in working memory. The integration of verbal and pictorial information is facilitated if text and graphics that belong together are presented together rather than separately (*spatial contiguity*) and animation and descriptive narration are played at the same time rather than one after the other (*temporal contiguity*) (Mayer 2005d). Practice exercises promote the integration of new information with existing knowledge because they require learners to connect what they know to what they are learning. Worked-out examples can be designed in such a way that problem-solving steps are gradually removed from an example and have to be supplemented by learners as they gain proficiency. This allows learners to move from studying prepared examples to completing partially worked-out examples to solving full-fledged problems on their own while being able to handle the cognitive demands of problem solving at each stage (Renkl 2005: 238f.).

4. *Retrieving and transferring long-term knowledge*. New knowledge and skills have to be encoded for storage in long-term memory in such a way that the learner can easily retrieve them back into working memory when he or she needs them later outside of the context in which they were acquired. Without successful retrieval, transfer of learned knowledge and skills does not occur. Accumulated and stored knowledge is of little use if learners are unable to apply this knowledge when and where they need it (Clark & Mayer 2003: 39f.). Thus, examples and activities should be appropriately contextualized by embedding them in scenarios that are similar to those in which learners will apply the acquired knowledge and skills later.
5. *Managing cognitive processes via metacognitive skills*. Since the resources available for cognitive processing are limited, learners have to select salient pieces from the incoming information and determine the degree to which connections should be created among the selected items and between new information and prior knowledge. *Metacognitive skills* help learners to allocate, monitor, coordinate, and adjust their limited cognitive resources during learning (Mayer 2005c: 36). These skills may be more or less pronounced in different learners and have to be developed (Clark & Mayer 2003: 40).

A major contribution of cognitive research on multimedia learning is a large number of principles for the design of multimedia learning environments, which have been derived from a substantial base of empirical research. These principles are discussed by Clark (2002), Clark and Mayer (2003), and by various authors in the *Cambridge Handbook of Multimedia Learning* (Mayer 2005a). Basic principles of multimedia learning described in these works are summarized in Table 1.

Cognitivism provides the theoretical basis for a second generation of computer-based learning software, which organizes and presents instructional content in a way that can be effectively processed by learners, given their inherent limitations. Instructional design is guided by cognitive principles, including those presented in Table 1. *Concept maps* (Novak & Cañas 2006) and other tools visualize the relationships between different concepts to facilitate the development of knowledge structures involving these concepts. Feedback from the computer goes beyond reinforcement and includes additional hints, comments, and information about potential sources of error (Arnold 2005: 8). Three further cognitivist principles for instructional design were described by Euler (1994). First, learners are given some control over the content, sequence, and pace of instruction, so they can study the material according to their own interests, needs, and preferences. Second, the instructional content is designed to be as close to reality as possible in order to make it easier for learners to integrate new information into their prior cognitive structures. Third, customized help is provided based on the learner's knowledge or performance level and progress. *Tutoring systems*, which include an instructional component in addition to practice exercises (cf. Section 2.3.2), and *intelligent tutoring systems*, which can tailor both instructional content and methods based on a model of the individual learner (cf. Section 2.3.3), are examples of second-generation educational software with a cognitivist background (Arnold 2005: 7f.).

Table 1. Principles of multimedia learning and their implications for instructional design.

Principle	Description	Implications
Multimedia principle (Fletcher & Tobias 2005)	People learn better from words augmented by pictures than from words presented alone. 'Words' may be printed or spoken words; 'pictures' can be static (e.g. diagrams and photos) or dynamic (e.g. animations and video clips) (Mayer 2005c: 32).	Augment the words of the instructional message with appropriate pictures that are relevant to the goal of instruction.
Split-attention principle (Ayres & Sweller 2005) (cf. contiguity principle)	People learn better when they do *not* have to split their attention between and devote cognitive resources to the integration of multiple information sources that are separate with respect to space or time, in order to make sense of the instructional message.	Words and pictures in instructional messages should be physically and temporally integrated rather than separated.
Modality principle (Low & Sweller 2005)	People learn better when information to be presented is meaningfully distributed among the visual and the auditory modes compared to when all information is presented in one of the two modes.	Present information using dual-mode rather than single-mode presentation techniques.
Redundancy principle (Sweller 2005b)	People learn better when multiple redundant forms of presenting the same information and unnecessary elaborations of information are avoided. Redundant information does not facilitate learning but makes it more difficult.	Eliminate all redundant materials and activities from multimedia lessons. Less may be more.
Contiguity principle (Mayer 2005d)	People learn better when corresponding verbal and pictorial information are presented together in space (*spatial contiguity*) or in time (*temporal contiguity*).	Present corresponding elements of a multimedia presentation together, both in space and in time.
Coherence principle (Mayer 2005d)	People learn better when a multimedia presentation does not contain extraneous (instructionally irrelevant) auditory and visual elements.	Avoid words, pictures, and sounds that are not directly relevant to the goal of instruction.
Signaling principle (Mayer 2005d)	People learn better when cues are added to a multimedia presentation that indicate what material is important.	Provide cues that guide the learner's attention toward important information.
Personalization principle (Moreno et al. 2001; Clark 2002; Clark & Mayer 2003: Chapter 8; Mayer 2005e)	People learn better when the language of a multimedia presentation is conversational (using *I* and *you* and direct comments to the learner) rather than formal and when pedagogical agents are available that provide instructional assistance (cf. Chapter 4).	Use conversational tone and/or pedagogical agents to increase learning by socially engaging the learner.
Voice principle (Mayer 2005e)	People learn better when a human voice with a standard accent (rather than one with a foreign accent, or a synthetic voice) speaks the words of a multimedia presentation.	Use a standard-accented human voice to deliver the words of a multimedia instructional message.
Image principle (Mayer 2005e)	People do *not* necessarily learn better when a multimedia presentation shows an image of the speaker (e.g. an animated agent) on the screen. However, the experimental evidence is inconclusive (Mayer 2005e: 208f.).	More research is needed to clarify the effects of using on-screen characters in multimedia lessons.

Other cognitivist approaches emphasize activities and tasks that require learners to apply problem-solving strategies (including search, trial and error, and exploration) (Schulmeister 1997: 71f.). Two important methods in this context include *learning by discovery* (Bruner 1961) and the use of *microworld simulations*, i.e. closed artificial environments representing selected aspects of the real world, which allow learners to perform experiments, try out different ways to solve problems, and develop a better understanding of the domain (Papert 1980; Romme 2002).

Cognitive approaches continue to make important contributions to educational psychology and the design of instructional systems (e.g. Clark & Mayer 2003; Clark & Lyons 2004; Mayer 2005a). Because it emphasizes the role of internal structures and processes of the mind, cognitivism can better account for complex forms of learning (reasoning, problem solving, and information processing) than behaviorist approaches (Ertmer & Newby 1993: 59). However, it has been criticized for focusing too much on the cognitive ('rational') processes in the human mind and neglecting the social, emotional, and motivational processes that influence learning, memory, thinking, etc. (Arnold 2005: 4).[6] Furthermore, cognitivism has retained the behaviorist views that knowledge exists independently of the individual (p. 4) and that learning processes can be controlled from the outside. Hence, instruction still has to find effective and efficient ways to communicate given knowledge to learners, which include simplification and standardization. Instructional content is sized, chunked, arranged, and weeded out so that learners can easily process the new information and incorporate it into their existing knowledge structures (Ertmer & Newby 1993: 59f.). All these aspects have been challenged by constructivist theories of learning.

2.2.3 Constructivism

During the 1980s and 1990s, a third paradigm gained a leading role in educational research, which grew out of dissatisfaction with the previously domineering views of learning through reinforcement (behaviorism) and learning through mental information processing (cognitivism). *Constructivism* (Duffy & Jonassen 1992; Ertmer & Newby 1993; Duffy & Cunningham 1996; Boethel & Dimock 1999; Arnold 2005) is an umbrella term for a broad spectrum of theories about knowledge and learning (Brooks & Brooks 1993: vii; Duffy & Cunningham 1996: 171) which share the view that knowledge is *not* given and absolute and mechanically acquired by learners (Liu & Matthews 2005: 387) but rather "a personal interpretation of experience" (Bednar et al. 1992: 21f.) that is actively and dynamically *constructed* by every individual for himself or herself through his or her experience in a particular context (Boethel & Dimock 1999: 6) and based on his or her prior knowledge, beliefs, and experiences. Like cognitivism, constructivism views learning as a mental activity (Ertmer & Newby 1993: 62). Learners are assumed to build an idiosyncratic internal representation of knowledge by processing new experiences in the context of their prior experiences and understandings, with the goal of establishing consistency (Boethel & Dimock 1999: 7). This internal knowledge structure can always be extended or modified based on new experiences (Ertmer & Newby 1993: 63).

6 However, the *cognitive-affective theory of learning with media (CATLM)* (Moreno 2005; Moreno & Mayer 2007), a recent expansion of the cognitive theory of multimedia learning, explicitly considers the mediating role of motivational and metacognitive factors in the learning process: the former increase or decrease cognitive engagement, whereas the latter regulate cognitive processing and affect (Moreno & Mayer 2007: 313f.).

Opportunities for learning arise when learners reach a state of *disequilibration* (also called *perturbation* or *puzzlement*) because they encounter something that is at odds with their previous understandings (Duffy & Cunningham 1996; Boethel & Dimock 1999: 7f.).

Constructivists reject the idea that knowledge can be transferred from one individual to another (e.g. from teacher to student). Instead, they emphasize that knowledge is the result of an individual active construction process and thus does not exist independently of the learner and the context in which it is constructed and used (Boethel & Dimock 1999: 6; Arnold 2005: 5). Knowledge is viewed as "a consensus of beliefs, a consensus open to continual negotiation" (Duffy & Cunningham 1996: 178).

In a constructivist learning environment, learners are required to actively deal with topics and tasks and make their own experiences, which amounts to learning by doing, exploring, playing, constructing, criticizing, etc. rather than by listening or reading alone (Schulmeister 1997: 74). The knowledge acquired in traditional classroom settings is mostly *inert knowledge*, i.e. students are taught information which they can reproduce in class, but which they fail to apply in real-world situations because they learned it out of context (Bednar et al. 1992: 23). Therefore, constructivism emphasizes that learning should be situated in a rich context that is as close to the real-world situation of learners as possible (p. 22). Learning is measured in terms of how effectively the learner's constructed knowledge representation facilitates his or her thinking and acting in the subject domain (Ertmer & Newby 1993: 64), not by counting the learner's correct answers or by assessing the correctness of his or her reasoning or problem solving.

Another important constructivist idea is that learning processes are embedded in social contexts and are facilitated and mediated by social interaction between learners and their peers and instructors (Dalgarno 1996; Boethel & Dimock 1999: 6). Finally, in constructivist settings, the role of the instructor changes from the "sage on the stage" (transmitter of knowledge) to the "guide on the side" (facilitator or coach), who provides opportunities, ideas, and help for meaningful learning (Issing 1995: 198; Boethel & Dimock 1999: 13). Knowledge and meaning are not prescribed by the instructor (which is pejoratively called *instructivism* by constructivists) but constructed by the individual learner or negotiated through social interaction among and between learners, instructors, and other parties (cf. Chapter 5.4.5). Hence, a major contribution of constructivism is that control over the process and outcome of learning shifts from the instructor to the individual learner or communities of learners. However, this shift does not diminish the role of the teacher, who in his or her capacity as instructional expert is still instrumental to successful learning, since computer-based tools (e.g. word processors, e-mail clients, and simulations) cannot provide meaningful interactions on their own (Boethel & Dimock 1999: 21). Cunningham summarized the constructivist position on instruction as follows:

> *The role of instruction in the constructivist view is to show students how to construct knowledge, to promote collaboration with others to show the multiple perspectives that can be brought to bear on a particular problem, and to arrive at self-chosen positions to which they can commit themselves, while realizing the basis of other views with which they may disagree. (Cunningham 1991: 14).*

There are two major variants of constructivism, which differ with respect to their view of the role of social interaction in learning, focusing on the individual learner and on learning as occurring in the context of social interaction, respectively (Boethel & Dimock 1999: 8). *Cognitive* or *radical* constructivism, which originates from the ideas of Jean Piaget (Piaget 1952; Piaget 1980), views learning as the result of an autonomous intrapersonal process of knowledge construction (Arnold 2005: 5), i.e. as the accomplishment of a single learner, who individually constructs knowledge as he or she learns (Duffy & Cunningham 1996). Social interaction is said to facilitate learning, but it is not regarded as essential. Radical constructivists focus on how the individual solves problems and constructs ideas (Boethel & Dimock 1999: 8) and hence stress the importance of learner-centered and discovery-oriented activities. In contrast, *social* or *realistic* constructivism, which is based on the work of Lev Vygotsky (Vygotsky 1962; Vygotsky et al. 1978), emphasizes the role of the social environment and regards learning as a social and cooperative process (Arnold 2005: 5) that is embedded in a learning community and in which learners construct knowledge through interaction with their immediate environment, in particular with their peers and instructors. From a social-constructivist point of view, language, culture, and context play important parts in the knowledge construction process (McMahon 1997; Liu & Matthews 2005). The most well-known concept of social constructivism is the *zone of proximal development* (Vygotsky et al. 1978), which is the distance between what learners can achieve on their own and what they can achieve through interaction with more capable others (Vygotsky et al. 1978: 86; McLoughlin & Oliver 1998: 128; [INT 14], cf. Chapter 12.3.2). With the help of more knowledgeable peers, instructors, or computer-based coaches, learners can master knowledge and skills which are beyond their individual capabilities. They are still responsible for actively constructing their own knowledge, but their interaction with facilitators is instrumental in the knowledge construction process and helps them to achieve a higher level of knowledge or proficiency (Cooper 2005: 31).

With its emphasis on social processes in learning, social constructivism is related to *situated learning* (Lave & Wenger 1991), a model of human learning which situates the acquisition of knowledge in the context of authentic social and physical environments ([INT 15]). Learning is viewed as a fundamentally social process in which the individual actively participates in collective learning taking place in social groups formed by practitioners (e.g. artists, engineers, nurses, etc.) who have competence in and commitment to a shared domain of interest and develop a shared practice by means of continuous interaction with each other over a longer period of time ([INT 16]). These *communities of practice* (Wenger 1998) are "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" ([INT 16]). The concept has found practical applications in various domains, including education, government, international development, organizational learning, and knowledge management ([INT 16]).

Another situated approach to learning is the *cognitive apprenticeship* model (Brown et al. 1989; Collins et al. 1991; Edmondson 2005: 19–22), which follows the example of the relationship between an apprentice and a master teacher, where the former learns a particular trade by working under the guidance of the latter in a realistic context (Edmondson 2005: 20f.). Likewise, in a cognitive apprenticeship, the master teacher (a human or a computer program, e.g. a pedagogical software agent, cf. Chapter 4) models concepts and skills for the learner in an authentic context. The learner observes the master as he or she is demonstrating the skill while explaining exactly what he or she is doing and why. These integrated performances and explanations help the learner to acquire relevant behaviors and a conceptual model of the skill (p. 21). In the next step, the learner practices the skill, imitating the master teacher (ideally both his or her behaviors and thinking processes), and the master provides *coaching* in the form of assistance (additional modeling, corrective feedback, encouragement, reminders, etc.) at those points which are just beyond the learner's capabilities (i.e. within his or her zone of proximal development) (p. 21). Over several repetitions of this process, the learner's skill level comes closer to the one of the master teacher, and the master gradually withdraws his or her support.

Eventually, the learner can perform the skill on his or her own, ideally at the same level of proficiency as the master (or at least close to it) (p. 21).

Constructivism is the currently predominant paradigm in educational research and has spawned a wide variety of theories, techniques, and systems related to learning. In the field of computer-based education, constructivist ideas have fundamentally changed both the view of the role of the machine in the learning process and approaches to the design of learning experiences (Young 2003). The task is no longer to teach pre-defined content and then tutor learners in it, based on goals and methods determined by the instructional designer in advance, but rather to create computer-based environments in which the learner can self-directedly engage in activities (building, criticizing, exploring, playing, writing, etc.) that promote individual knowledge construction processes, as well as interact with other learners or human and computer-based coaches (Arnold 2005: 10). These self-directed activities are embedded in authentic contexts; are meaningful to learners; focus on the application, not the acquisition of knowledge; and exhibit realistic complexity rather than simplification for instructional purposes (Boethel & Dimock 1999: 14). "We must *maintain* the complexity of the environment and help the student to understand the concept embedded in the multiple complex environments in which it is found" (Bednar et al. 1992: 26, their emphasis). Computers and other media are not primarily delivery systems for instructional content (although some instruction may be involved where appropriate); instead, they function as *cognitive tools* (Lajoie & Derry 1993; Jonassen & Reeves 1996; Jonassen et al. 1998; Lajoie 2000) for the active construction of knowledge (Arnold 2005: 10). Cognitive tools may range in complexity from ordinary word processors to sophisticated simulation systems (p. 20, cf. Section 2.3.4). To facilitate social interaction, communication tools such as e-mail, discussion forums, chat rooms, instant messaging, and VoIP ("Voice over IP," Internet telephony) may be provided. In general, in constructivist learning environments, the computer supports but does not direct or control the learning process, unlike (intelligent) tutoring systems (cf. Section 2.3.2 and Section 2.3.3).

A number of broad guidelines for the design of learning experiences can be derived from constructivist principles. They include (Boethel & Dimock 1999: 12–15; Arnold 2005: 10f.):

- Create authentic learning environments providing meaningful contexts for activities.
- Start by introducing a problem rather than by explaining concepts, theories, or facts.
- Relate problems and topics to learners' lives, knowledge, and experiences.
- Prefer complex, ill-structured problems over tasks simplified for didactic purposes.
- Make room and provide the tools for active knowledge construction processes.
- Give learners as much freedom of choice and exploration as possible.
- Provide help and guidance where necessary (but without exercising control).
- Present the learning matter in different contexts and from multiple perspectives.
- Aim for in-depth coverage of central ideas rather than superficial treatment of many topics.
- Promote communication and cooperation with other learners and with instructors.
- Involve learners in a cognitive apprenticeship or in a community of practice.
- Encourage learners' articulation of and reflection on their knowledge and strategies.
- Regard students' errors as opportunities for learning. Guide learners through puzzlement.
- Assess learners' ability to transfer learning to new problems and situations.
- Involve the target group of learners in the design process (cf. Chapter 7.13).

Constructivist ideas have unquestionably made a significant impact on the ways in which people think about knowledge and learning processes. The merits of constructivism include its view of learning as an active process of knowledge construction, which is individual and idiosyncratic but at the same time has a strong social component, and its doing away with the idea that learning processes can be controlled from the outside (Arnold 2005: 6). While these are common themes, it is also true that there is no single constructivist theory but a whole family of partially overlapping, partially competing theories can be subsumed under that label. In fact, “because there are so many versions of constructivism, with important overlaps, but also with major differences, it is difficult to see the forest for the trees” (Phillips 1995: 7). As a result, when some approach or design is labeled as ‘constructivist,’ it is necessary to determine which constructivist framework underlies it in order to understand the exact implications of the label in that context.[7] The list of guidelines above can help to determine the constructivist orientation of a given computer-based learning scenario, but an instructional design that follows (some of) these guidelines is not necessarily constructivist in nature and even may not be intended to be. On the other hand, when the label ‘constructivist’ is claimed for a design, there is no guarantee that constructivist design guidelines have actually been applied (Arnold 2005: 14). Because constructivism is the current mainstream, people are quick to adorn their course materials and systems with that label, but more often than not these materials or systems hardly follow any of the constructivist design guidelines listed above (p. 14).

A more philosophical criticism of constructivist approaches accuses them of tending toward *relativism* (e.g. Phillips 1995), i.e. the belief that truth is not absolute but always relative to a language, a culture, or some other referential framework ([INT 17]). In response to this criticism, Bereiter noted with respect to the constructivist assertion at issue that “there is no objective standpoint” (Boethel & Dimock 1999: 10):

> *This statement is sometimes taken simplistically to mean that there are no absolute truths, but that misses the point. The point is that there is no objective standpoint from which to judge whether something is an absolute truth. Also, the statement does not mean that there is no real, material world. It only means that we can never get down to an objective knowledge of it that can serve as a basis for our judgments, beliefs, and interpretations (Bereiter 1994: 4).*

Likewise, constructivist claims regarding the subjectivity of the learner’s knowledge and the idiosyncratic nature of his or her view of reality (self-constructed through learning) have given rise to some concern. In general, it is not advisable for any theory of knowledge and learning to risk losing the connection with the real world because the real world imposes considerable constraints on human knowledge construction (Phillips 1995: 12; Stone & Goodyear 1995). Furthermore, there are situations in which conformity may be important and idiosyncratic action or thinking may cause problems. For example, Merrill asked, “Do we want students to have a ‘self-chosen position’ with regard to the sound of letters in learning to read? Do we want students to have a ‘self-chosen position’ about the meaning of integers?” (Merrill 1992: 107). Or is it wise to allow learners to construct their own perspective on traffic rules when learning how to drive a car? The controversy is about whether there is knowledge that is ‘objective’ and can (perhaps even should) be taught to learners, or whether all conceptions of learners are necessarily idiosyncratic. Constructivists’ response to the criticism by Merrill and others is that learners inevitably develop their own position with regard to any subject matter (Boethel & Dimock 1999: 10). They argue that the construction of exclusively idiosyncratic conceptions (Duit 1995: 274) is prevented through social interactions that negotiate a shared understanding which enables coexistence and communication with others (Jonassen 1994). Still, for instructional designers and practitioners, the basic organizational problem remains that they

[7] The same is true of the other two paradigms. Behaviorism and cognitivism are also both umbrella terms for families of different but related theories.

have to find a balance “between supporting students in reaching their own understandings and steering them toward an accepted body of knowledge, i.e. the required curriculum” (Boethel & Dimock 1999: 12). “Teachers necessarily want to encourage students to make constructions that are personally meaningful, and yet recognize that they must also construct ideas that are acceptable to the wider society” (Wood 1995: 337). Another problem is the quasi-religious and ideological side of constructivism (Liu & Matthews 2005: 386): for many theorists, their particular version of constructivism has turned into a “secular religion,” and non-believers or members of other constructivist ‘sects’ are looked (down) upon with distrust (Phillips 1995: 5).

2.2.4 Conclusion

Despite its dominance, constructivism has not managed to oust behaviorism and cognitivism completely. In fact, both earlier paradigms continue to play an important role in educational theory and practice. While constructivists insist that “the findings of constructivism *replace* rather than add to our current understanding of learning” (Bednar et al. 1992: 30, emphasis added), constructivism has not brought the “end of instruction” (Kerres 1998), and most likely never will. Each of the three approaches, behaviorism, cognitivism, and constructivism, has its own niche in the design of effective learning experiences. Behaviorist principles form the basis of designs that reinforce simple skills and recollection of facts (know-what). Cognitivism informs the design of instructional experiences that equip learners with the know-how to solve problems by transferring learned knowledge and skills to unfamiliar situations (Ertmer & Newby 1993: 68). In addition, cognitive research has contributed valuable guidelines for the design of multimedia learning environments that facilitate the mental processing of information (cf. Table 1). Constructivist ideas have inspired a shift in the focus of instructional design from the externally controlled transfer of knowledge toward learning experiences that emphasize learner control, cooperation, authenticity, reflection, and multiple perspectives. The constructivist approach is particularly suitable for domains involving complex and ill-defined tasks and problems that require a high level of cognitive processing (Ertmer & Newby 1993: 68). Actively participating in cognitive apprenticeships and communities of practice can help people to learn how to deal successfully with problems and tasks of this kind.

It is interesting to note that some constructivists seem to agree that behaviorist and cognitivist approaches have a place in learning processes beside constructivism. For example, Jonassen (1991) distinguished between three stages of knowledge acquisition (introductory, advanced, and expert) and acknowledged that beginning learners may benefit more from objectivist (i.e. behaviorist and/or cognitivist) approaches during the introductory stage of knowledge acquisition. As learners become more advanced, they are better equipped to handle the complex and ill-structured problems presented in constructivist learning environments, which help them to discover, negotiate, and possibly modify or remove misconceptions and biases from the introductory stage (Ertmer & Newby 1993: 65). It appears that eclectism rather than dogmatism is the appropriate strategy for instructional design given the various theories of learning and their different practical implications. There is no single ‘best’ approach that works for every learner, subject matter, situation, competence level, and stage of learning. Instead, the task of the instructional designer is to pick those methods that achieve optimal instructional outcomes in the current design context (Ertmer & Newby 1993: 69f.).

2.3 E-Learning Technologies

This section reviews a number of computer-based technologies that are used to create e-learning experiences. The review is selective rather than comprehensive and focuses on technologies that are important for this thesis.

2.3.1 Drill-and-Practice Systems

Drill-and-practice systems provide environments for the structured reinforcement (cf. Section 2.2.1) of previously acquired knowledge and skills (Rist & Hewer 1996). Learners are assumed to have some background in the subject matter and are presented with questions and exercises giving them the opportunity to practice or recapitulate what they have learned (Grob & Seufert 1996). The activities may be of varying levels of difficulty and may be selected by the system at random or in sequence, for example. Normally, learners have to successfully complete tasks at a given level before they are allowed to proceed to the next higher level. Their answers are analyzed (typically in terms of 'correct' vs. 'incorrect') and generate immediate feedback from the system (Pellone 1995; Arnold 2005: 7). Both the learner's correct and incorrect answers are stored. Incorrect answers or failure to pass an entire task result in the repeated presentation of the respective question or task until the learner gives the correct answer or passes the task (Pellone 1995). Interactive and multimedia elements are often used to enrich drill-and-practice activities, like for example in the Interactive Tutor facility of the Virtual Linguistics Campus (cf. Section 2.4.1). Furthermore, the questions and exercises can be embedded in a game scenario that provides attractive rewards for the successful completion of tasks (Rist & Hewer 1996, cf. Section 2.3.5).

2.3.2 Tutoring Systems

Tutoring systems are computer-based instructional systems that involve the computer in the role of an impersonal tutor teaching new topics in a given subject area, usually to a single learner. A tutoring system is equipped with knowledge about both the content of instruction and strategies for teaching it to learners. In addition, it typically also provides an environment in which learners can practice what they have learned. Hence, tutoring systems are like drill-and-practice programs in some ways, but they additionally have an instructional component (Arnold 2005: 7). "The instructional emphasis is on displaying material to the learner, checking for comprehension by monitoring student behavior on narrow tasks with readily observable outcomes, then branching to further presentation or practice as appropriate" (Dede & Lewis 1995: 10). While allowing for some degree of learner control (over the content, sequence, and pace) (Arnold 2005: 7), the tutor is firmly in charge of the learning process, "programming learners through a graduated sequence of skills and concepts" (Dede & Lewis 1995: 10). Three broad categories of tutoring system can be distinguished (Schulmeister 1997: 96f.; Meier 2006: 154ff.). *Linear* tutoring systems only provide a single path of instruction, which the learner cannot leave. These programs are also called *page turners* because learners can only move backward and forward, similar to turning the pages of a book. *Branching* sequences of instruction are more flexible because they offer different paths of instruction depending on user inputs, leading the learner to alternative presentations, additional explanations, or questions at different levels of difficulty. In both linear and branching models, the possible paths of instruction are wired into the system. In contrast, *flexible* or *open* program structures dynamically generate instructional paths, depending on the duration of the session, the progress of learners, their degree of motivation, or their requests. This requires sophisticated learner modeling capabilities, possibly based on techniques from artificial intelligence (see the next section).

2.3.3 Intelligent Tutoring Systems

Tutoring systems generally lack the ability to adapt the content and method of the instruction to the individual learner. *Intelligent tutoring systems (ITS)* (Wenger 1987; Shute & Psotka 1996; Freedman 2000; Padayachee 2002) have been developed to address this problem. Using techniques developed in the field of artificial intelligence (AI), an ITS attempts to emulate the skills of a good private human tutor (Servan-Schreiber 1987), which include the ability to:

- Provide information;
- Solve problems independently and explain solutions;
- Analyze and diagnose the learner's behaviors and mistakes;
- Determine the learner's understanding of the topic;
- Adapt subject matters and teaching methods to the learner;
- Communicate effectively with the learner.

The four main components of an intelligent tutoring system are shown in Figure 5 (e.g. Wenger 1987: Chapter 2; Schulmeister 1997: 181–191; Freedman 2000; Padayachee 2002). The *domain model* organizes knowledge about the subject matter to be taught. The *student model* represents the system's evolving view of the learner's understanding of the subject matter. It is developed by diagnosing and recording the learner's responses, and using the accumulating evidence to infer what the learner knows, does not know, and which aspects of his or her knowledge are incorrect. The *tutor model* embodies the teaching expertise of the ITS. It helps to design the content, form, and timing of interventions from the system in ways that are intended to benefit the learner. The domain model, the student model, and the tutor (or instructor) model will be discussed in detail in Chapter 12.1 to Chapter 12.3. Finally, the *tutor interface* provides input and output devices for meaningful interaction between learner and ITS, possibly involving natural language dialogue. Techniques for processing human-machine conversations in natural language will be discussed in Chapter 11.

2.3.4 Simulations

An educational simulation is a closed artificial environment providing a dynamic model of a (real or fictitious) process, system, or situation, which can be influenced by the learner's actions or decisions prior to or while running the simulation. Simulations may receive a continuous stream of inputs (including actions and sensory percepts) from the learner and respond in real time, or the learner may set certain parameters first and then run the simulation to see the effects of his or her manipulation (Bostock 1995). Learners can observe simulations from the outside or, using, for example, virtual reality technology, become *immersed* in them, i.e. a part of the simulation. When a simulation represents aspects of reality, its *fidelity* becomes important, i.e. how closely the simulation models the original process, system, or situation.

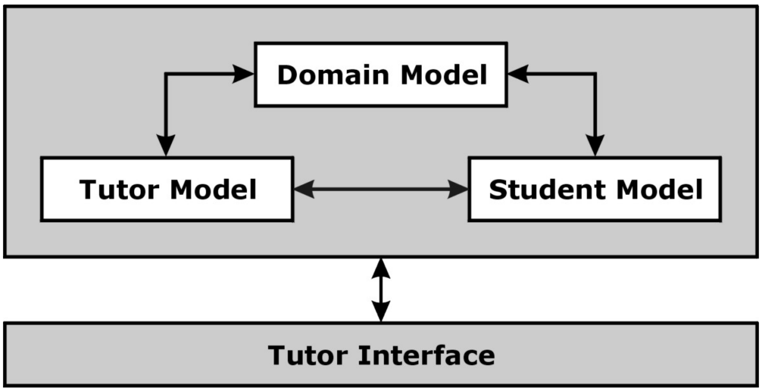

Figure 5. The architecture of an intelligent tutoring system.

Simulations create virtual learning environments (sometimes called *microworlds*) that serve as contexts for active, experiential learning in domains including the natural and social sciences, economics, and the operation of vehicles and machinery (Schulmeister 1997: 67). In these environments, feedback is intrinsic to the simulation rather than provided explicitly (as in tutoring systems) (Bostock 1995). Likewise, "much of the knowledge to be gained is implicit and contextual, mastered by the student through reflection on the results of interactions with the synthetic environment" (Dede & Lewis 1995: 13).

One advantage of simulations is that they provide relatively risk-free environments for exploration and training (Mason & Rennie 2006: 106). The effects of learners' actions and decisions (including those of his or her mistakes) are confined to the simulated environment and therefore cause no real harm, loss, or damage, neither to the learner nor to other people, objects, or equipment. Furthermore, simulations may allow learners to do or experience things that are impossible or too dangerous in real life (e.g. swimming in the blood stream through the body or walking around the core of a nuclear reactor). Finally, educational simulations can serve as a replacement for the real thing (e.g. the dissection of animals).

2.3.5 Educational Games

Computer games continue to be popular with people from all ages, and recently, there has been a surge of interest in online multiplayer games. Combined with the fact that "today's learning generation is extremely game literate" (DeKanter 2005: 28), this has encouraged exploration of the potential of educational games as an e-learning activity (Mason & Rennie 2006: 53f.). For example, Purushotma examined "how content originally designed for entertainment purposes can be modified to provide natural and context-rich language learning environments, without sacrificing its entertainment value" (Purushotma 2005: 80).

Playing educational games is said to increase learners' motivation and to promote acquisition of knowledge and skills in the course of game play. Learning occurs by doing things (alone or together with others) in the game context, sometimes without learners being aware of it (Mason & Rennie 2006: 54). Game environments can be as simple as a game-show-like question-answer format or as complex as a three-dimensional virtual reality simulation (p. 54) in which real and virtual humans (embodied software agents) can interact with each other (e.g. Johnson et al. 2005a).

Effective educational games continuously present the learner with different challenges and offer motivating rewards, which makes them particularly attractive for younger groups of learners (Mason & Rennie 2006: 54), although the average age of computer game players is increasing (Castronova 2006: 57). Games can allow learners to play various roles in both single-player and multi-player settings. Likewise, the computer and other learners can play different parts in the game experience, such as opponent, ally, referee, coach, etc. (cf. Chapter 3.2.4). Educational multiplayer game simulations can promote collaborative learning of teams in authentic, challenging scenarios from areas such as business, government, or military (e.g. Rickel & Johnson 2003; Bonk & Dennen 2005).

2.3.6 Hypertext and Hypermedia

Coined by Theodor Nelson (Nelson 1965), the terms *hypertext* and *hypermedia* are used to refer to a non-linear organization of objects, which consists of a network established by means of connections (so-called *hyperlinks*) between related, self-contained units of information (Kuhlen 1991: 27; Tergan 1997: 6). Both hypertext and hypermedia are closely connected to the World Wide Web.

By following hyperlinks, a particular information unit can be reached from any number of other units linking to it. Unlike the linear presentations found in many textbooks, the users of hypertext or hypermedia systems are not restricted to following a single path through the presentation of material but can browse through the network, forming associations between linked pieces of information according to their own interests (Dede & Lewis 1995: 11). The difference between hypertext and hypermedia lies in the contents of the nodes of the network. While the information units in a hypertext system consist of text, hypermedia systems support different types of media (images, audio, video, etc.) in nodes in addition to text (Tergan 1997: 6). However, for most authors, the terms ‘hypertext’ and ‘hypermedia’ are interchangeable ([INT 18]). For example, Landow wrote:

> *Because hypertext systems link together passages of verbal text with images as easily as they link two or more passages of text, hypertext includes hypermedia, and I therefore use the two terms interchangeably. (Landow 2006: 84).*

The associative nature of hypertext and hypermedia is believed to be more effective for learning than linear presentations because it is assumed to be closer to the way in which information is processed by the human mind (Shapiro & Niederhauser 2004: 608; [INT 18]): information about one topic usually evokes associations with other pieces of information about the same topic or different topics (Faber 1993). The resulting networks of associations may be more important than knowledge of isolated facts because they enable learners to understand patterns of relationships (Dede & Lewis 1995: 11f.). The linked structure of the information space in hypertext or hypermedia systems is claimed to facilitate this kind of associative processing because it gives learners immediate access to the information that is associated with a node (Faber 1993). Having said that, it can be difficult for people to locate specific pieces of information in complex hypertext or hypermedia structures and to navigate these spaces without losing their way in the web of cross-references and experiencing information overload because of the excess of available links ([INT 18]). This points to the need for effective navigation and search tools in hypertext or hypermedia environments. However, the changing habits and practices of people with respect to ingesting and digesting information also have to be taken into account. While previous generations were schooled in the reading and comprehension of linear text ([INT 18]), the Net Generation obtains much of its information from the World Wide Web, which is the most well-known implementation of the hypertext/hypermedia concept ([INT

19]; [INT 20]). Hence, it may be more natural for younger learners to deal with hypertext or hypermedia structures than for their parents or instructors.

2.3.7 Multimedia

Multimedia (Tannenbaum 1998) is "the combination of various digital media types, such as text, images, sound, and video, into an integrated multi-sensory interactive application or presentation to convey a message or information to an audience" (Neo & Neo 2001: 20). Multimedia learning environments deliver content to learners over multiple sensory channels simultaneously, ideally in ways that observe the capabilities and limitations of human cognition, as outlined in Section 2.2.2. However, multimedia not only provides vivid presentations of instructional material. In fact, a major feature of multimedia systems is their interactivity[8] (Neo & Neo 2001: 20), which allows learners to manipulate elements of a multimedia presentation to produce certain illustrative and/or entertaining effects (in fact, instruction and entertainment are not incompatible in multimedia systems but symbiotically feed off each other). Another quality of multimedia content is that it may be linear or non-linear ([INT 21]). Learners can experience different versions of a non-linear multimedia presentation by choosing among multiple paths through the content according to their interests (Dede & Lewis 1995: 11). In hypermedia systems, learners have considerable freedom to associate different pieces of multimedia information connected through hyperlinks (cf. Section 2.3.6). Finally, multimedia systems can be designed to accommodate users with different learning styles by allowing them to choose combinations of media that suit their individual preferences (auditory, visual, or symbolic) (p. 11).

'Multimedia' is a term that was very popular in the mid-1990s but has come to be used to describe so many different experiences that it has lost much of its descriptive value (Mason & Rennie 2006: 82). More than ten years later, multimedia is no longer a novel concept; in fact, it is something that is more or less expected from state-of-the-art environments for learning, entertainment, and information.

2.3.8 Computer-Mediated Communication

Computer-mediated communication (CMC) (Thurlow et al. 2004) is an umbrella term for various technologies that support human-human communication over computer networks, including e-mail, chat, discussion forums, newsgroups, instant messaging, Internet telephony, audio and video conferencing. The role of technology is to mediate communications between two or more individuals rather than to provide or process information (Bostock 1995). CMC can be *synchronous* (the communication occurs in real time with all participants online, which applies to chat, instant messaging, Internet telephony, audio and video conferencing) or *asynchronous* (senders and recipients of messages are temporally decoupled; examples include e-mail, discussion forums, and newsgroups) and may involve sender-recipient relations of one-to-one, one-to-many, or many-to-many (Bostock 1995). Most CMC today still happens through

[8] It is worth noting here that interactivity is a concept in want of a precise and widely accepted definition. The term is ubiquitous, and people seem to have some intuitive idea of what it means, but the concept is actually very hard to pin down. Different ways of conceptualizing interactivity will be discussed in Chapter 5.

text, although the use of other media, such as audio and video, is becoming more and more important.

2.3.9 E-Learning 2.0

E-Learning 2.0 (Downes 2005; Karrer 2007) is a current buzzword used to refer collectively to a “second generation” of e-learning that is based on the philosophy and technologies of *Web 2.0*. This is another collective term for various recent social and technological developments that are changing how people experience and use the World Wide Web (O’Reilly 2005; Panke 2007). While there is still no precise definition of the concept, for many, Web 2.0 materializes as ‘social’ web-based applications that are easy to use, invite (actually rely on) user participation and collaboration, and promote the formation of communities and social networks among users. Important goals of Web 2.0 include increasing the involvement of users in the creation of content, improving the reusability of content, and making the search for information more effective (O’Reilly 2005; Panke 2007: 3). Often-cited examples of Web 2.0 applications are Flickr ([INT 22]), a site for publishing and sharing photographs, del.icio.us ([INT 23]), a site for sharing bookmarks, and Wikipedia ([INT 24]), a multilingual, web-based, free content encyclopedia editable by anyone ([INT 25]). Web 2.0 is commonly associated with a number of technologies, including:

- *Weblogs* (or *blogs*) are web applications that provide simple content management features for creating, editing, deleting, organizing, and displaying regular time-stamped contributions (posts) written by a single author or (sometimes) a group of authors (Mason & Rennie 2006: 14f.; Panke 2007: 17). Blog posts are shown in their temporal sequence on a common web page, typically starting with the latest post (Mason & Rennie 2006: 14). They contain their authors’ comments and thoughts on particular topics; however, weblogs go beyond personal journals in at least three ways: the posts may contain embedded images, video clips, etc. and links to other web resources; the contents of a blog is accessible to anyone; and most weblogs invite public comments on their posts from users and thus may gradually gain their own communities of readers (p. 15). A recent trend in blogging involves the creation of *video blogs (vlogs)*, i.e. blogs whose entries consist (primarily) of video clips.
- *Web syndication* is a form of content distribution that allows web sites to use content made available by other web sites ([INT 26]). News and weblog sites were among the first to use web syndication, but today, all kinds of digital information are distributed in this way. Web syndication commonly involves publishers providing frequently updated lists of content from their sites (so-called *web feeds*), and consumers who subscribe to these feeds in order to obtain new syndicated content for display in their own sites or browsers, automatically or on demand ([INT 26]). A web feed ([INT 27]) is a structured document written in one member of the RSS (Rich Site Summary or Really Simple Syndication) family of formalisms, which are based on XML (eXtensible Markup Language) (cf. Footnote 182).
- *Podcasting* (derived from the blend of *iPod* and *broadcast*) is a syndication-based method by which audio (and video) content can be automatically distributed over the Internet to subscribers for playback on their portable media player or computer, allowing them to access the downloaded content when (and where) it is convenient for them ([INT 28]). Users subscribe to a *podcast* (i.e. a set of digital media files) syndicated (made available) by its author or host (the *podcaster*) and automatically receive new content when it becomes available ([INT 28]).

- *Tagging* is the human activity of associating a particular piece of information (e.g. a photo, blog entry, map, or video clip) with freely chosen descriptive keywords (*tags*). Users of sites like del.icio.us and Flickr add tags to information in order to make it easier for themselves and for others to find the tagged information later (Panke 2007: 16). As a result of this collaborative categorization effort by a large number of people, user-created taxonomies of metadata emerge, so-called *folksonomies* (cf. Chapter 12.1.2.5) (p. 16).
- *Wikis* are web-based content management systems that enable collaborative authoring on a large scale by allowing users to add, modify, and remove (almost) any content, including the content created by others, easily and quickly using a web browser. The name comes from Hawaiian *wiki wiki*, which means 'quick' (Mason & Rennie 2006: 130) and indicates the ease and speed with which content can be created and refined in a wiki system. Wikis differ from other groupware systems by the concepts of open and collaborative editing, community ownership of content, and constant flux (or change) (pp. 130f.). Furthermore, wikis rely on social mechanisms established by the community of users rather than on strict workflows, access restrictions, and formal structures enforced by the software to regulate collaborative editing processes and maintain order (p. 131).

The emphasis of Web 2.0 on interactivity, participation, collaboration, and social networking has drawn the attention of both researchers and practitioners in the field of e-learning because these characteristics are compatible with (social-) constructivist views of learning (cf. Section 2.2.3) (e.g. Brown & Adler 2008). It is believed that Web 2.0 technologies can be used to give learners an active, participatory role in the learning process by making them authors of learning content, which they create from their own point of view and using information from all kinds of sources collected from the World Wide Web. Often, the content created by learners will reflect their own interests more than topics of the course (Downes 2005). During the authoring process, learners can develop skills in information literacy, critical thinking, writing, and self-expression (Mason & Rennie 2006: 15). Furthermore, for example, in the course of writing one blog entry after the other or of making successive changes to a wiki document, learners may be encouraged to reflect on their thinking and writing over a longer period of time (p. 15). Through the sharing of blog entries and comments, the collaborative creation of web documents in a wiki, or cooperation to develop a taxonomy for the information in a domain through tagging, social networks and communities can form between learners, which are similar to Wenger's communities of practice (Downes 2005, cf. Section 2.2.3). In all these scenarios, the role of e-learning changes from medium to platform, from a tool for content delivery to a tool for content authoring (Downes 2005). In addition, E-Learning 2.0 has implications for the distribution of learning content. The traditional format of the pre-packaged complete online course is giving way to the syndication of *microcontent* (i.e. content conveying a single idea or concept in a form that is succinct and suitable for access and presentation through e-mail clients, web browsers, or mobile devices, cf. Dash 2002) from different sources, which learners aggregate, remix, and repurpose according to their interests and needs (Downes 2005; MacManus & Porter 2005).

2.3.10 E-Learning Platforms

An e-learning platform is a software system that can host one or more e-learning environments and provides services for managing the content and users of these environments. Depending on their focus, two major types of e-learning platforms can be distinguished: *learning management* *systems*, which emphasize the automation of administrative tasks and *content management systems*, which are tools for authoring and organizing e-learning content.

2.3.10.1 Learning Management Systems

The central concern of a learning management system (LMS)[9] is the automation of selected aspects of the administration of the e-learning process. Typically, this includes the management of content, resources, instructors, and students. Content authoring capabilities are normally not part of a learning management system; instead, LMSs focus on delivering and managing content created by external sources.

For the most part, learning management systems have been implemented as web-based applications to give learners, instructors, and administrators location- and time-independent access to learning materials and administrative tools ([INT 29]). Elements of a typical LMS include facilities for ([INT 29]):

- Assembling courses from learning objects;
- Handling roster operations (learner accounts and enrollment in courses);
- Managing learners, instructors, courses, resources, roles, access, and privileges;
- Calendaring and scheduling of activities, deadlines, events, etc.;
- Sending notifications and other messages to learners;
- Communication among learners, instructors, and administrative staff;
- Tracking, assessing, and reporting on learner's performance and progress;
- Pre-testing and post-testing of learner's knowledge and skills;
- Grading course work and generating scores and transcripts;
- Supporting both online and blended course delivery (cf. Section 2.1 and Section 2.4.2).

Although a wide range of both commercial and open source learning management systems are available today, quite a number of successful e-learning systems, including the Virtual Linguistics Campus (cf. Section 2.4), have chosen to build their own LMS. The decision to make one's own rather than buy or adopt an external learning management system is motivated by several disadvantages of the latter:

- The LMS may be difficult to use due to an excess of features, confusing interface, bugs, etc.
- The LMS may be incompatible with one's e-learning concept and philosophy.
- The LMS may include features that are not needed and may miss essential others.
- The LMS may be difficult to customize (corporate identity, additional functions, etc.).
- The LMS may not integrate easily with existing components and systems.
- The LMS may involve high initial and follow-up costs (customization, licensing, etc.).
- The LMS may cease to be developed and supported some day.

9 A number of other, similar terms are used to describe essentially the same functionality (with a few nuances): virtual learning environment (VLE), course management system (CMS), managed learning environment (MLE), learning support system (LSS), and learning platform (LP), among others (Mason & Rennie 2006: xxxiii; [INT 30]).

2.3.10.2 Content Management Systems

A content management system (CMS) is a web-based or other computer application designed to organize and support all activities involved in the collaborative process of designing, testing, revising, approving, and publishing documents and other content which are typically part of a larger structured collection, such as a web site ([INT 31]). Web-based CMSs are often used to manage the different contents of large web sites ([INT 32]).

In contrast to LMSs, content management systems provide dedicated facilities for authoring content, including the possibility to define *templates*, i.e. layouts with placeholders for *assets* (text, images, and other media), which can be filled in from a library of assets, by authors, or automatically on the fly, in order to create presentations of content. Modern CMSs usually include a database back-end which facilitates the storage and reuse of assets and content created using these assets.

Content management systems for e-learning applications are often called *learning content management systems (LCMSs)*. An LCMS allows both experienced and less experienced users to create, manage, combine, and present *learning objects*, i.e. reusable, self-contained units of learning (Mason & Rennie 2006: 71), in order to create personalized e-learning experiences ([INT 29]).

However, the use of (L)CMSs also has a number of disadvantages. These systems often generate static, cookie-cutter layouts of images and text that lack the flexibility and originality of designs developed by multimedia experts. Another problem of (L)CMSs, LMSs, and e-learning platforms in general is that they can lure people, in particular administrators, into thinking that simply buying and installing such a system will suffice to create an e-learning culture in their organization. In other words, e-learning requires much more than a sophisticated software product.

2.4 The Virtual Linguistics Campus

The *Virtual Linguistics Campus (VLC)* ([INT 3]) is the leading platform for e-learning in linguistics worldwide (Handke & Franke 2006). The VLC is the result of a multidisciplinary research and development effort since 2001, which involved teams from the universities of Duisburg-Essen, Marburg, and Wuppertal in the first phase (2001–2003) and was then carried on by the Linguistic Engineering Team (LET) at the University of Marburg (2003–today). A screenshot of the VLC homepage in October 2007 is shown in Figure 6.

After several years of development, today, a wealth of certified courses and study material in the fields of theoretical linguistics, applied linguistics, linguistic engineering, and teacher training is available from the Virtual Linguistics Campus, supported by various tools for administration, communication, and linguistic education ([INT 3]). All VLC courses are suitable for both online and blended learning. A third option is to use selected materials from e-learning units of the VLC to supplement face-to-face instruction (see Section 2.4.2 for a discussion of these scenarios). The target group of the VLC includes students, instructors, theoretical and applied linguists, school teachers, university or college administrators, and other interested parties.

Since 2001, the Virtual Linguistics Campus has been an integral part of regular teaching activities at various universities in Germany and other countries. In October 2007, the VLC had over 1,200 registered users from more than 55 countries, and the rising trend continues. The majority of users are students of language and/or linguistics in their early-to-mid twenties, of

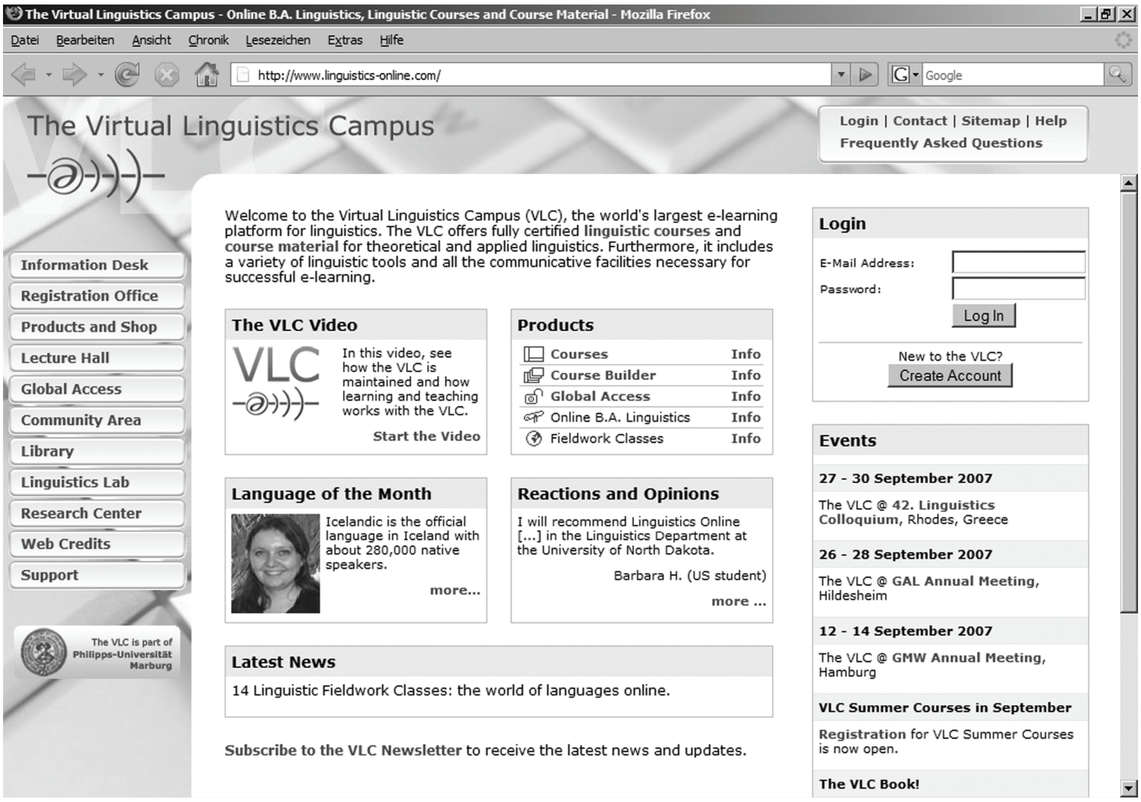

Figure 6. The VLC homepage in October 2007. Screenshot taken on 20 October 2007.

both sexes (although there are significantly more female than male students), and mostly from Western (in particular German) culture.

The VLC has recently spawned two further e-learning platforms: the *Virtual Center for Teacher Training (VCT)* ([INT 33]) and the *LET-Vokabeltrainer* ("LET Vocabulary Trainer") ([INT 34]). The VCT offers certified online courses for teachers and students of language and literature on e-learning, web development, content presentation techniques, and the new media in (foreign-language) education. The LET-Vokabeltrainer provides various exercises designed to train students from German schools in the vocabulary of (currently) English and Latin that they are required to learn as part of their curriculum. Each of the three platforms has a different target group of users: the VLC is designed for university students, the VCT for students and professionals of teaching, and the LET-Vokabeltrainer for learners in schools.

Section 2.4.1 below outlines the structure and components of the VLC (for a detailed description, see Handke 2006a). The following section discusses how the VLC can be used to support different e-learning scenarios, including face-to-face teaching (cf. Section 2.4.2.1), online learning (cf. Section 2.4.2.2), and blended learning (cf. Section 2.4.2.3). Section 2.4.3 describes the linguistic fieldwork classes on the VLC, which provide the setting for the different pedagogical agent roles discussed in Chapter 16.

2.4.1 The Architecture of the VLC

The Virtual Linguistics Campus is a web-based e-learning platform that provides an interactive learning environment integrating multimedia content, educational and recreational activities, information resources, and tools for administration, assessment, communication, research, community building and learner support. Where possible, the platform makes use of multimedia and database technologies to provide dynamic and interactive activities and content.

The system architecture of the VLC is shown in Figure 7. It consists of three major components: a ‘free,’ unrestricted area, a ‘secure’ area whose contents is accessible only to (authorized) logged-in users of the campus, and an administration component for use by the VLC staff only. The free area provides general information about the VLC, what it has to offer, and how to use the platform. Furthermore, this area gives access to help facilities, demos, linguistic tools, and various other information and services that may be useful or interesting for visitors (Handke 2006a: 15). For a discussion of the options available in the free area, see Handke (2006a: 15ff.). However, most of the contents of the VLC are in the secure, restricted area (cf. Figure 7). To gain access to this part, people have to become registered users of the VLC first by creating a user account. With their account, they can then enroll in classes on the VLC or subscribe to the Global Access facility (see below). After successful enrollment or subscription, they can access the corresponding protected parts of the VLC through their personal page (“My VLC”). The third, administrative component became necessary to facilitate previously manual and labor-intensive management tasks in the constantly changing and growing environment of the VLC (Handke 2006a: 30), which include class compilation and administration, user management, and registration procedures, among others. The *VLC Management System* (Handke 2006a: 30–33) models the data structures required for the administration of the VLC as a set of databases and provides a web-based interface for managing modules, courses, classes, e-learning units, users, dates and deadlines, topic suggestions for term papers, and so on. The purpose of this interface is to hide the complexity of the underlying structure from non-developers (p. 32, cf. Chapter 1.2).

The Virtual Linguistics Campus is first and foremost a platform which provides instructional content to learners, both as stand-alone units and as part of online courses. All courses on the VLC are designed in such a way that learners can study them completely online or blended, i.e. with supplemental in-class teaching (Handke 2006a: 20, cf. Section 2.4.2).

The e-learning concept of the VLC is based on the following hierarchy (Handke 2006a: 20 + 31f.). The *Lecture Hall* lists all VLC courses, grouped under their modules. Each *module* contains the VLC courses for a certain broad field of study (currently including theoretical linguistics, applied linguistics, teacher training, and linguistic fieldwork). Within a module, learners study *courses* (academic units) for credit by first registering for and then participating in one of their classes. Examples of VLC courses include Phonetics, Human Language Technologies, English Linguistics (for Teachers), and the linguistic fieldwork course Hindi. The *classes* within a course may differ (slightly) with respect to their content, the institution offering them via the VLC, and the person(s) teaching them. A given class has a fixed time period during which it is active and participants have access to the *e-learning units* of the class. Figure 8 shows the structure of a typical VLC e-learning unit, which consists of the following components (Handke 2006a: 23f. + 27f.):

- The *Introduction* screen lists the learning goals and central topics of the unit, provides a definition (if appropriate), specifies a set of bibliographical references for further reading, and includes a link to specific advice for learners about studying the unit.
- The *Virtual Session* is the core of an e-learning unit. Being the digital equivalent of a conventional lesson, it presents the instructional content, but in a multimedia and interactive format that emphasizes appropriate use of vivid graphics, sound, video, and animations over showing text-heavy pages (Handke 2006a: 28). The basic principle of the Virtual Sessions is that of explorative learning, guided by a set of questions that cover the important aspects of a Virtual Session (p. 27). Through the menu bar of the Virtual Session, learners can access a

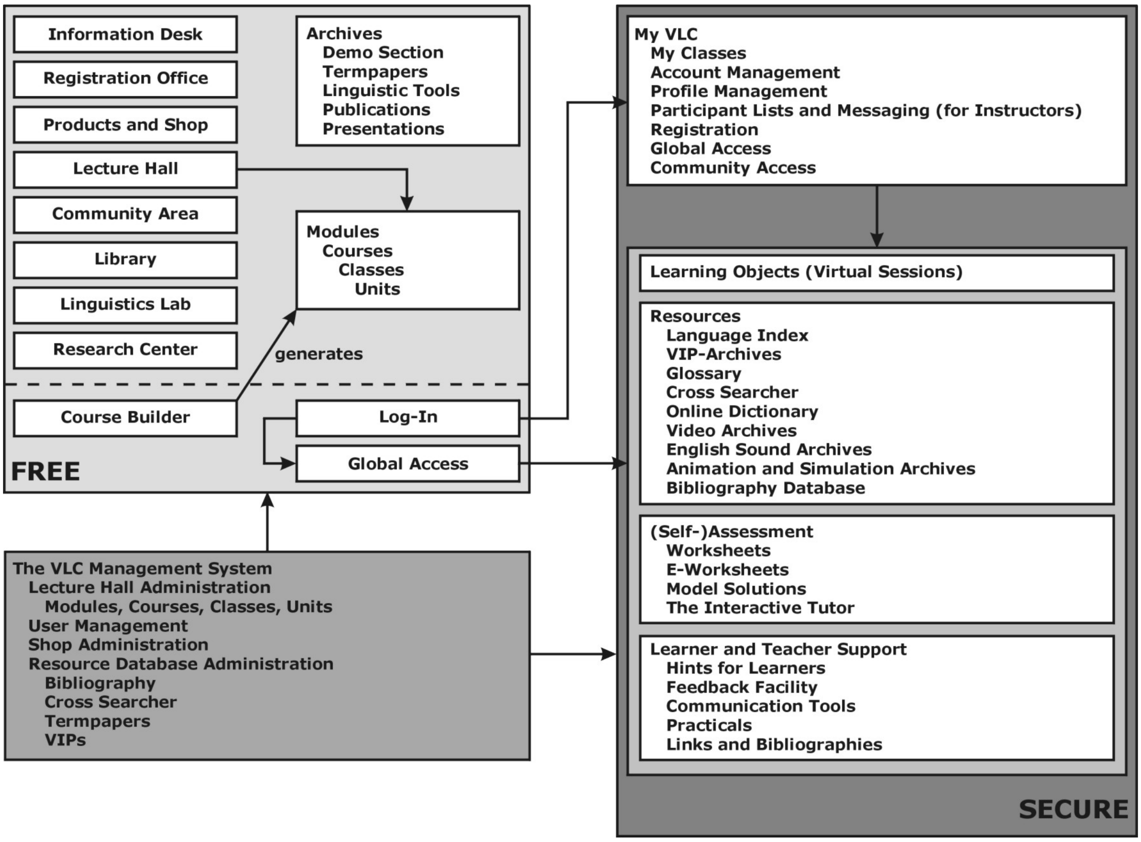

Figure 7. The architecture of the VLC. Adapted from Handke (2006a: 14, Figure 1.1).

Glossary of linguistic terms, display information about linguistic concepts from other Virtual Sessions via the *Cross Searcher*, search a *Language Index* presenting multimedia information about more than 75 languages, and read the biographies of key figures in the field of linguistics in the *VIP-Archives* (pp. 28f.). After studying the Virtual Session, learners can practice/reinforce what they have learned using the *Interactive Tutor*. This component provides various kinds of interactive exercises, including multiple-choice questions, listening tasks, transcription tasks, drag-and-drop exercises, identification tasks, input tasks, etc.

- The *Worksheet* is a collection of linguistic analysis tasks, which learners typically have to submit as part of their credit requirements. Worksheets are implemented as web-based forms and enriched with multimedia elements where appropriate for a task. Students can submit them individually or in small groups of up to three people. Each Worksheet has a submission deadline. After the deadline, the model solution to the Worksheet is made available. The *E-Worksheet* is a variant of the Worksheet, which consists exclusively of online exercises that are generated, corrected, and evaluated by the computer. Both the instructor and the learner receive a notification after the learner has passed an E-Worksheet (Hente 2006: 76f.).
- The *Practical* consists of a collection of exercises, problems, and tasks related to the topics of the unit, which is intended mainly as a resource for instructors of blended classes (cf. Section 2.4.2.3) who want to enrich their in-class sessions with additional activities. Hence, the Practical is like the traditional handout in some respects (Franke et al. 2006: 9), but unlike the handout, it may contain embedded multimedia elements. Model solutions to the tasks are available, which makes the Practical a valuable resource for both blended learners and online learners (who can try to solve the Practical tasks on their own and then look at the solutions) (Handke 2006a: 24).

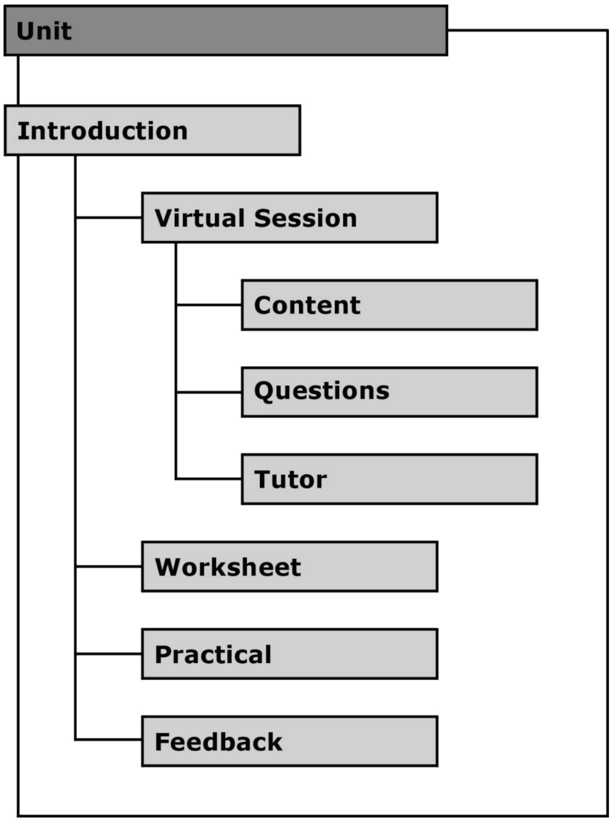

Figure 8. The structure of an e-learning unit on the VLC. Adapted from Handke (2006a: 23, Figure 1.8).

- The *Feedback* facility is a web-based form that learners can use to submit questions or comments to the relevant contact person (e.g. questions about the Virtual Session to the instructor, technical inquiries to the webmaster, etc.).

Booking and attending classes is the typical way of using the VLC. However, for those learners who want to study the Virtual Sessions without registering for classes (e.g. when they are preparing for an exam), the *Global Access* facility provides class-independent access to the Virtual Sessions of the VLC (Handke 2006a: 18). If teachers, universities, or other institutions are interested in building their own linguistic e-learning classes, they can compile them from the units provided by the VLC using the *Course Builder* (p. 19). These custom classes are delivered via the VLC and supervised by the person or institution that ordered them.

The *VLC Community Network* is a recent addition to the Virtual Linguistics Campus. This feature gives users of the VLC the possibility to learn about other users, create friendship contacts, and exchange messages with their friends. To become a part of this network, users have to make their profile visible to others and set the visibility options for the different items in their user profile, which include a variety of personal information such as nickname, country, and e-mail address.

Recently, a printed *Workbook* has been created for each course to serve as a study companion for learners (Unger 2006a). The Workbook comprises the contents of the Virtual Sessions in a serialized (rather than hypertextual) format, with many blank spaces for the learner to fill in while studying the Virtual Sessions, as well as pointers to online interactive multimedia elements. In addition, for each unit, the Workbook provides an introduction to the content, central goals, and topics; the Practical; and the bibliographical references.

2.4.2 E-Learning Scenarios Involving the VLC

The structure and contents of the Virtual Linguistics Campus facilitate its use in three instructional settings that involve different degrees of face-to-face contact, dependency on technology, and use of online components: face-to-face (f2f), online, and blended. The basic relationship between these scenarios is shown in Figure 9. In the first setting, selected elements of the VLC are used to support face-to-face teaching. In the second scenario, learners are studying VLC courses completely online, without any in-class teaching. Between these two extremes, blended learning courses based on the VLC combine the e-learning units of the campus with in-class practical sessions with the instructor, which focus on discussion and problem solving rather than on teaching of content. These three settings are discussed in the following subsection, using material that was previously published elsewhere (Franke 2003).

2.4.2.1 Support for Face-to-Face Teaching

Some instructors may want to integrate individual features of the Virtual Linguistics Campus into their regular face-to-face teaching but may not wish to change the basic approach and structure of their courses. These instructors can use selected elements of the VLC where they deem it appropriate in order to enrich their in-class teaching or to provide additional activities and resources for their students. Examples of how the VLC can support face-to-face teaching include:

- Illustration of linguistic concepts with animations (e.g. articulation of speech sounds);
- Discussions in the chat room or discussion forum;
- Student assessment using the Interactive Tutor and (E-) Worksheets;
- Preparatory and follow-up studying of Virtual Sessions for an in-class lecture;
- Replacement of (e.g. canceled) in-class lectures with e-learning units.

Hence, components of the VLC are used in an eclectic fashion to contribute to what is still essentially face-to-face, in-class instruction. E-learning elements play a supporting rather than a primary role. This approach is useful for introducing interested teachers to e-learning without immediately confronting them with the full range of tasks involved in managing a technology-based blended or online course.

2.4.2.2 Online Learning

Online courses on the Virtual Linguistics Campus are delivered entirely through the VLC-platform. They are characterized by the primarily technology-based interaction of learners with multimedia instructional content, instructors, and other learners. Learners can study the e-learning units of an online course using a standard web browser in their own time and at their own pace. Apart from constitutional meetings and in-class exams in some courses, there are no face-to-face meetings during an online course. Learners communicate with their instructors and peers through e-mail, chat rooms, discussion forums, or (rarely) over the telephone. Students in VLC-based online courses often come from different countries all over the world, but online learning has also become an attractive option for learners enrolled at the University of Marburg, the home of the VLC, to help them to resolve conflicts between simultaneous mandatory courses in their schedule.

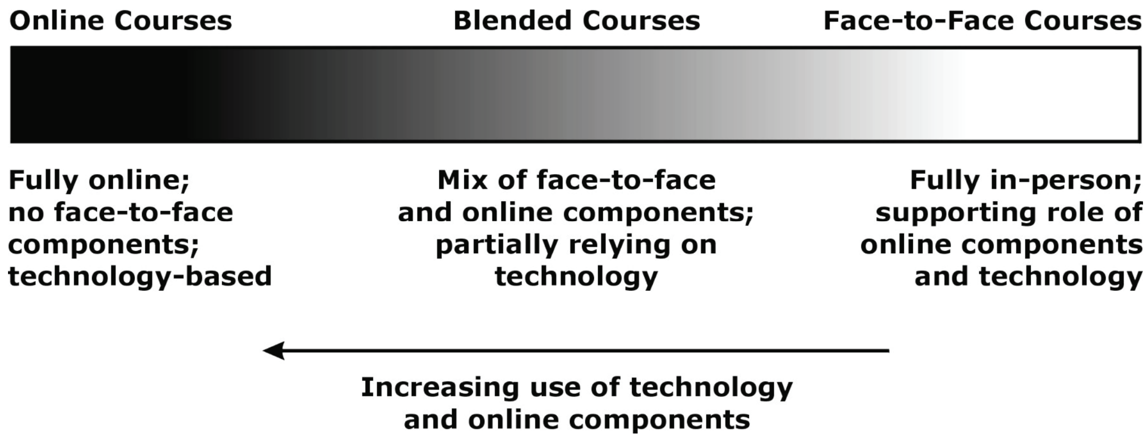

Figure 9. Online, blended, and face-to-face (f2f) courses compared. Adapted from Mason and Rennie (2006: 14, Figure 4).

For each e-learning unit in an online course, learners study the Virtual Session first, including the Interactive Tutor exercises, and then optionally submit the (E-) Worksheet of the unit as part of their credit requirements. In addition, they may post in the discussion forum, participate in chat sessions, and make friends and exchange messages with their peers in the VLC Community Network. The nature of the online learning situation (distributed learners, no face-to-face interaction with instructors and peers) and its potential negative effects on learners (uncertainty, isolation, frustration, decline of motivation, and eventually dropout) makes constant communication a necessity in order to establish and maintain contact and relationships between all individuals involved in an online course. The experience with online courses on the VLC has shown that the course instructor has to take the initiative in contact cultivation. The following is a list of qualities of an online instructor that have positive effects on the motivation of the individual learner and the overall atmosphere in an online course (see, for example, Kemshal-Bell (2001) and Denis et al. (2004) for a detailed discussion):

- Good availability (e-mail, chat, telephone, in person, etc.);
- Timely and detailed feedback on learners' questions, solutions, performance, etc.;
- Ability to solve learners' technical problems;
- Advisory capabilities (course choice, time management, learning strategies, etc.);
- Source of knowledge, insight, and inspiration;
- Help with the development of metacognitive skills;
- "Tender love and care" for each learner, in particular the passive and frustrated ones;
- Skills in facilitating online discussion and collaboration;
- Promotion of learner-instructor and learner-learner relationships;
- Competent course management (record keeping, scheduling, enrollment, etc.).

Overall, the role of the instructor in an online course changes from transmitter of knowledge (this is largely the responsibility of the technology) to facilitator of learning processes (at the cognitive, administrative, and technological levels) (Denis et al. 2004) as well as of social relationships between the participants.

Online courses as outlined above are offered regularly on the Virtual Linguistics Campus. Students appreciate the flexibility and constant availability which this format provides. However, the lack of face-to-face contact with the instructor and other learners is a problem for online learners. The blended course format combines the advantages of both face-to-face teaching and online courses.

2.4.2.3 Blended Learning

> *Year after year, face-to-face interactions are ranked by all students in either first or second place. This replicates the results of many distance education studies that show students often feel that something important to their learning is missing when all interactions are mediated, whether asynchronous or synchronous. (Dede 2004: 7).*

As social beings, humans tend to feel uncomfortable when they are out of contact with other people for longer periods of time. We need to see, hear, and feel others around us; in other words, we crave human contact. The human need to socialize with others is an explanation for their strong preference for learning experiences that involve face-to-face interaction with instructors and other students and their uneasiness in online courses, where they normally do not meet their instructors or peers in person (with the exception of kick-off meetings and in-class examinations) (Guri-Rosenblit 2005: 12).

Both face-to-face teaching and online learning have their strengths and weaknesses (cf. Table 2). While online learning was originally propagated as an alternative to face-to-face instruction, the two approaches have recently come to be seen as complementary rather than competitive, and have been integrated into a hybrid approach to the organization of instruction. The term *blended* (or *hybrid*) *learning* gained popularity around 2000 and commonly labels combinations of online learning and face-to-face teaching (although other blends of technologies, resources, methods, and locations have also been labeled “blended learning”) (Mason & Rennie 2006: 11f.).

The basic idea of blended learning is to achieve synergetic effects in learning by combining the strengths of online courses with those of face-to-face teaching. The mix of online and face-to-face elements is intended to give learners as many options as possible in order to suit their individual learning needs and learning styles. Recent research indicates that blended learning is more effective and more enjoyable for learners than either face-to-face teaching or online learning alone, and preferred by instructors and learners alike (Dean et al. 2001; Aycock et al. 2002; Guri-Rosenblit 2005; Rossett et al. 2003; Rovai & Jordan 2004; Frankle 2005). The challenge for instructional design is to develop a mixture of online and face-to-face components which maximizes success in learning (Kerres & de Witt 2003).

Blended learning on the Virtual Linguistics Campus builds on the online e-learning units and combines them with face-to-face sessions with the instructor in the classroom. All blended courses on the VLC follow a common scheme:

1. Learners study the Virtual Session online at a place, time, and pace that suits them.
2. Learners optionally submit the (E-) Worksheet of the e-learning unit.
3. Learners participate in weekly in-class practical sessions with the instructor.

The major difference between online courses and blended courses on the VLC lies in the additional face-to-face sessions with the instructor, which take place every week during a semester. Unlike traditional face-to-face lessons, the purpose of the in-class practical sessions is *not* to teach content. Instead, their focus is on giving learners the opportunity to discuss and practice what they have learned during their online study of the Virtual Sessions. Building on the background provided by the current Virtual Session, the instructor can use the in-class time flexibly to deepen students’ understanding of the topics discussed in the Virtual Session through various activities and exercises. In general, the practical sessions require careful preparation on the part of the instructor. Not only should the instructor be familiar with the structure and contents of the Virtual Session, but he or she should also come to the class equipped with meaningful activities based on the topics of the Virtual Session that illustrate important concepts, introduce new perspectives, give learners practice in linguistic analysis and other skills, and so on. The Practical of the e-learning unit provides valuable resources for the design of the in-class session (cf. Section 2.4.1). Furthermore, the instructor should be open to learners' questions, suggestions, and needs as they arise in the course of the session. There is plenty of room for this, since the instructor does not have to get through a certain amount of instructional content as in a traditional lesson. Because of the different focus of the in-class practical sessions, the instructor should emphasize discussion, practice, and analysis over presentation (Hente 2006: 79) and try to overcome the passiveness of learners common in traditional class formats by involving all attendants in stimulating activities, giving learners the opportunity to actively participate in the learning experience by asking questions, contributing to discussions, performing analyses, etc. The experience with blended learning courses on the Virtual Linguistics Campus supports the accumulating evidence (see above) that blended learning is an effective instructional format, which is well-received by learners and instructors.

2.4.3 Explorative Learning on the VLC

The focus of many courses on the Virtual Linguistics Campus is on delivering instructional content about a particular subject area to learners, using interactive multimedia elements and various kinds of (online and in-class) exercises and other activities to illustrate linguistic concepts and deepen learners' understanding of the topics discussed. In contrast, the linguistic fieldwork classes on the VLC (Unger 2006b) take a different approach. Simulating the classical linguistic field situation, they confront learners with an unknown language and provide them with the facilities to explore the language and discover as many details about it as possible. These facilities include (Unger 2006b):

- A virtual native speaker playing the role of informant;
- A virtual room featuring the virtual native speaker and various objects, which the speaker may simply label or use in more complex constructions when the learner clicks on them;
- General information about the language;
- Background on how to do linguistic fieldwork;
- Summaries of current hypotheses about the language based on the data analyzed so far;
- Growing vocabulary lists for the language;
- Various linguistic tools (e.g. sound charts, help with phonetic transcription, etc.).

Linguistic fieldwork on the VLC is a process that, on the one hand, challenges the learner and encourages him or her to engage in the active construction of knowledge (cf. Section 2.2.3), both about the target language and about linguistic analysis methods in general but on the other hand can quickly overwhelm especially beginning learners with its complexity. For the field research to succeed, it is necessary to inform, guide, and facilitate learner's exploration of the target language. In real-world fieldwork scenarios, support from different human actors may be available, including native speakers, coaches, and peers. With the exception of the virtual native speaker, these actors have no virtual counterpart in the fieldwork classes on the VLC so far. Chapter 16 will describe virtual actors (so-called pedagogical software agents, cf. Chapter 4) playing these roles to assist learners in the fieldwork situation.

Table 2. The advantages and disadvantages (marked grey) of face-to-face teaching and online learning.

I. Face-to-Face Teaching	II. Online Learning
The face-to-face environment facilitates social contact, relationship building, group formation, and instructional conversations.	Learners have greater flexibility. They can work in their own time and at their own pace. They can access the online learning environment at their own convenience, anywhere and anytime.
Instructional conversations can flexibly shift between instructor-learner, learner-learner and multi-party formats.	Online instructional material can be easily updated and exhibits constant quality (compared to the variable performance of human instructors).
The participants naturally coordinate verbal and non-verbal communication channels to negotiate meaning and resolve ambiguity.	The combined use of multiple media (images, audio, video, animations, and text) accommodates different types of learners.
Both learners and instructors can flexibly ask questions (including follow-up questions) and receive feedback immediately.	Learners have simple and instantaneous access to a wide range of distributed information resources.
Instructional conversations can be directed by the instructor toward desired learning outcomes.	The online learning environment can be designed to adapt to the learner's abilities, preferences, and progress, based on constant monitoring and modeling of different aspects of the learner.
Learners can easily socialize with their peers and instructors.	Learners can become immersed (i.e. experience presence) in simulated environments, manipulate them, and immediately observe the effects. These environments allow for the safe and inexpensive training of complex skills.
Instructors can adapt their teaching flexibly to learners' needs, for example in reaction to questions, suggestions, and perceived or pronounced issues.	Computers are infinitely 'patient' and 'friendly.' Feedback from computers is immediate, objective, and positive/non-judgmental.
Face-to-face instructional settings involve a familiar and comprehensible allocation of roles, which is in line with sociocultural stereotypes about education ("the instructor teaches learners; students learn from their teacher").	Learners can collaborate with real and virtual partners to accomplish tasks, using both synchronous and asynchronous computer-mediated communication technologies.
Instructors and learners have to be in the same place at the same time. Learners who have several parallel courses or other commitments may find it difficult to resolve the conflicts in their schedule.	Both learners and instructors fear a loss of social interaction between and the resulting isolation of the participants in online courses.
Differences between learners in terms of relevant previous knowledge and abilities can hinder the learning progress of the entire group.	Possible incorrect conceptions of learners arising from their study of the online material can remain undetected by the online instructor because there is no direct feedback from learners.
The weaker students in the group determine the pace of instruction, to the disadvantage of the more capable learners.	Online learners have to motivate and discipline themselves considerably, which they may not be prepared for.
In-class lessons are no longer available once they are over, except for handouts, notes, slides, etc. This can be a problem for learners who missed lessons or want to recapitulate them.	Self-directed learners without guidance may lose orientation, feel overwhelmed by the wealth of information, and fail to achieve learning goals.

2.5 Summary

This chapter reviewed selected concepts, theories, and technologies related to e-learning that are relevant to the present work. Furthermore, it introduced the Virtual Linguistics Campus, the e-learning environment that will become the 'playground' of the embodied pedagogical software agents whose design is the concern of this thesis.

The goal of Section 2.1 was to clarify, by means of an elaborate definition, what the term 'e-learning' is taken to imply in the context of the present research. In addition, the relationship of e-learning to other, related terms, including face-to-face instruction, online learning, blended learning, mobile learning, and distance education was discussed in this section.

The second section focused on learning and instruction, defining and contrasting the two and distinguishing instruction from education, training, and teaching. Instructional design (ID) was discussed as a process that builds a bridge between theoretical research on human learning and instructional practice, by translating the results of the former into plans of action for the latter (Ertmer & Newby 1993: 50). Three influential theoretical perspectives on human learning and their implications for instructional design were reviewed: learning through reinforcement (behaviorism), learning through mental information processing (cognitivism), and learning through individual construction of knowledge (constructivism). It was concluded that each of the three approaches has its own niche in instructional design and that designers should be eclectic (rather than dogmatic) with respect to their choice of methods based on the different approaches.

Technologies involved in the creation of e-learning experiences were discussed in the third section. The review covered technologies which enable learners to practice knowledge and skills (cf. Section 2.3.1); teach instructional content in a structured manner (cf. Section 2.3.2); can adapt instruction to the individual learner (cf. Section 2.3.3); provide simulated experiences for learners (cf. Section 2.3.4); engage learners in games with educational goals (cf. Section 2.3.5); present content as hyperlinked networks of self-contained units of information (cf. Section 2.3.6); combine multiple media for multisensory delivery of instructional content (cf. Section 2.3.7); provide tools for mediating communication between learners over computer networks (cf. Section 2.3.8); allow learners to author their own content and form social networks (cf. Section 2.3.9); and, finally, facilitate the management and authoring of e-learning experiences (cf. Section 2.3.10).

Section 2.4 focused on the Virtual Linguistics Campus. After summarizing its architecture, components, and features, it was shown how the VLC is used in three instructional settings: face-to-face teaching, online learning, and blended learning. The section concluded with a description of explorative learning on the VLC, as it occurs in the linguistic fieldwork classes. Three roles for virtual actors (i.e. pedagogical agents) in the fieldwork scenario were identified: the virtual native speaker, the coach, and the peer. These roles will be discussed in detail in Chapter 16.

3 Software Agents

Information and communication technologies are always increasing not only in sophistication but also in complexity while their *usability* (i.e. how effectively, efficiently, and satisfactorily people can use them to achieve their goals, cf. Chapter 6.3) has not substantially improved. Computers are still largely passive recipients of specific, highly detailed instructions whose execution needs to be monitored by the user. Generally, the user is expected to adapt to the machine while the reverse would be desirable to improve the accessibility of computers for the ever increasing number of non-expert users. In particular, "the interaction between computer and user (…) must become an equal partnership – the machine should not just act as a dumb receptor of task descriptions but should *cooperate* with the user to achieve their goal" (Jennings & Wooldridge 1998b: 7, their emphasis).

Intelligent software agents (Genesereth & Ketchpel 1994; Wooldridge & Jennings 1995; Nwana 1996; Jennings & Wooldridge 1996; Bradshaw 1997; Jennings et al. 1998; Wagner 2000; Russell & Norvig 2003) are claimed to solve (or at least mitigate) these problems by offering informed, adaptive, autonomous, and proactive assistance to users and other agents. Particularly at the user interface (cf. Chapter 1.2), software agents can thus transform the interaction from user-controlled command execution into a cooperative process (Maes 1994: 31).

In very general terms, an *agent* is an entity that perceives its environment and acts on it in some way (Russell & Norvig 2003: 32, cf. Figure 10). The agent is said to be *situated* in its environment (Jennings et al. 1998: 8). Agents perceive their environment through *sensors* and act on it by means of *actuators* (or *effectors*) (Russell & Norvig 2003: 32). Examples of sensors of artificial agents include keyboard, mouse, microphone, camera, accelerometer, temperature probe, Global Positioning System (GPS), etc. Some actuators include loudspeaker, display, virtual or mechanic body (parts), brakes, heater, lights, etc.

This definition of agency is intentionally broad, encompassing such diverse entities as thermostats (and other kinds of control systems), animals, humans, robots,[10] software daemons,[11] non-player characters in computer games, and virtual peers and tutors. The environments in which agents are embedded may be physical or virtual and usually contain objects and other agents (both natural and artificial). Sensor inputs include percepts of objects, states, events, and relationships as well as communications from other agents.

'Intelligent' agents have the capacity for autonomous, flexible, and goal-directed interaction with their environment, for themselves or on behalf of other agents. They are able to detect relevant environmental states and to perform actions in the environment in pursuit of their own agendas to achieve desired results (Rosenschein 2001: 411).[12] While *natural* agents (humans

---

[10] A *robot* is a "machine that senses, thinks, and acts" (Bekey 2005: 2). Robots completely consist of electromechanical components (Menzel & D'Aluisio 2000; Craig 2004; Bekey 2005).

[11] A *software daemon* is a process (program being executed on a computer) that monitors a software environment (often the computer's operating system) from the background and carries out operations to manipulate this environment according to a pre-defined schedule or when specific events occur.

[12] Whereas traditional approaches to building intelligent machines viewed intelligence as a property of "disembodied minds," more recent embodied and situational approaches to machine intelligence (Pfeifer & Scheier 1999) emphasize that intelligent behavior critically involves the body and its sensory-motor processes (Chrisley 2004: 62). "Intelligence is viewed as the capacity for real-time, situated activity, typically inseparable from and often fully interleaved with perception and action.

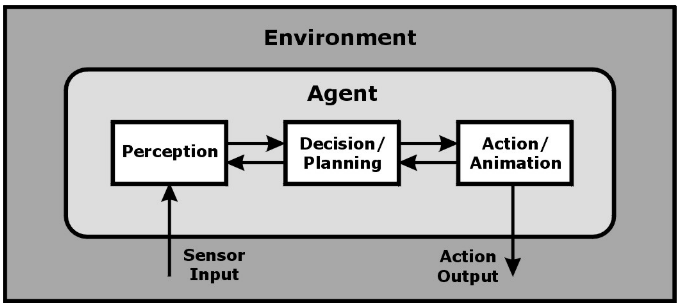

Figure 10. An agent situated in its environment. Adapted from Russell and Norvig (2003: 33, Figure 2.1). The agent is equipped with sensors through which it perceives its environment, and produces as output actions (including animations) that affect it, mediated by some kind of decision-making planning stage (Russell & Norvig 2003: 32f.).

and animals) exhibit all of these characteristics to varying degrees, the intelligence of *artificial* (human-made) agents is still inferior in comparison to that of natural agents.

Intelligent software agents are software systems loosely conforming to the above criteria, which inhabit (are situated in) a computer or computer network environment, learn the habits and interests of human users, and can operate autonomously on users' behalf to accomplish tasks delegated to them (Maes 1994; Maes 1995). The classic image of a software agent is that of a personal assistant (Maes 1994; Negroponte 1995: Chapter 12) with these qualities (cf. Section 3.2.5). Properties of intelligent software agents are discussed in the following section.

3.1 Properties of Intelligent Software Agents

Software agents are known by many other names, which emphasize particular characteristics or functions. The following is an incomplete list:

- Autonomous agents;
- Intelligent agents;
- Adaptive interfaces;
- Virtual assistants;
- Softbots;
- Entertainment bots;
- File sharing bots;
- Shopping bots;
- E-mail agents;
- Network agents;
- Printing agents;
- Search agents;
- Knowledge navigators;
- Guides.

The variety of labels indicates that agent research and development cover an increasingly wide range of agent types. It has even been argued that the term 'agent' is overused these days because everybody seems to call everything an agent (Nwana 1996), depending on how broad the scope of their definition of the term is, in many cases to promote a product by giving it a stylish label. Quite often, the difference between agents and 'ordinary' computer programs is not clear (Franklin & Graesser 1996).

Further, it is by having a body that a system is situated in the world in order to perform tasks that might previously have been thought to require the manipulation of internal representations or data structures" (p. 62). See Chapter 14.1.2 for a discussion of agent architectures based on this approach.

Although the term 'agent' is pervasive, perhaps even overly so, the notion of agency to date has defied attempts to produce a single definition that would satisfy everybody (Wooldridge 1999: 28), the one provided in this chapter being just one attempt among many others. Still, it is possible to identify a set of properties of intelligent software agents (Wooldridge & Jennings 1995: 116ff.) which many agent researchers would accept and which different types of agents possess to a greater or lesser extent. These properties also distinguish software agents from conventional programs: "[a]ll software agents are programs, but not all programs are agents" (Franklin & Graesser 1996).

3.1.1 Autonomy

Autonomy means that an agent has the ability to act in a complex and dynamic environment for an extended period of time without (constantly) requiring direct instructions from humans or other agents (Wooldridge & Jennings 1995: 116). Figure 11 shows a taxonomy of different kinds of autonomous agents (Franklin & Graesser 1996), including *biological agents* (humans and most animals), *robotic agents* (cyborgs[13] and mobile robots), and *computational agents* (autonomous agents that inhabit computers and networks). The latter can be subdivided into *software agents* (see above) and *artificial life agents* (autonomous agents that are born, grow, live, reproduce, die, and evolve in computer-simulated artificial environments)[14] (Franklin & Graesser 1996).

An autonomous agent is in control of its actions and internal state (Wooldridge & Jennings 1995: 116), and acts to achieve a set of goals, which may be its own goals or those of others. This implies that the agent can decide by itself on the basis of its sensor input and internal state whether to perform a certain action. The internal state includes what the agent currently holds true about its environment, other agents, and itself (*beliefs*); what it would like (and could attempt) to achieve (*desires* or *goals*); and what it has opted to attempt to achieve (*intentions*) ([INT 35], cf. Chapter 14.1.1).

Fully autonomous agents require minimum human intervention, which gives them maximum flexibility but can also make them unpredictable and raises issues of their trustworthiness because the agent may have goals of its own which do not necessarily serve the human user. Semi-autonomous agents, in contrast, are less flexible because they require a higher degree of human intervention, but they also give their users a greater sense of being in control. From this point of view, users should be able to control the degree of agent autonomy (Sycara et al. 1996), preferably at the task level.

3.1.2 Adaptiveness

The ability to acquire, change, and improve their behaviors as they react to and/or interact with their environment is critical for intelligent agents. An intelligent software agent can learn from

---

[13] The term *cyborg* is an abbreviation for "*cyb*ernetic *org*anism," which denotes a hybrid of organic and machine (synthetic) components contained in a (human) body, with the synthetic parts added to enhance the body's natural abilities (Gray et al. 1995; Mann & Niedzviecki 2001; [INT 59]).

[14] The various chapters in the volume edited by Langton (1997) provide an overview of the field of artificial life. A well-known artificial life simulator is the Creatures system (Grand et al. 1997; Grand & Cliff 1998; Cliff & Grand 1999).

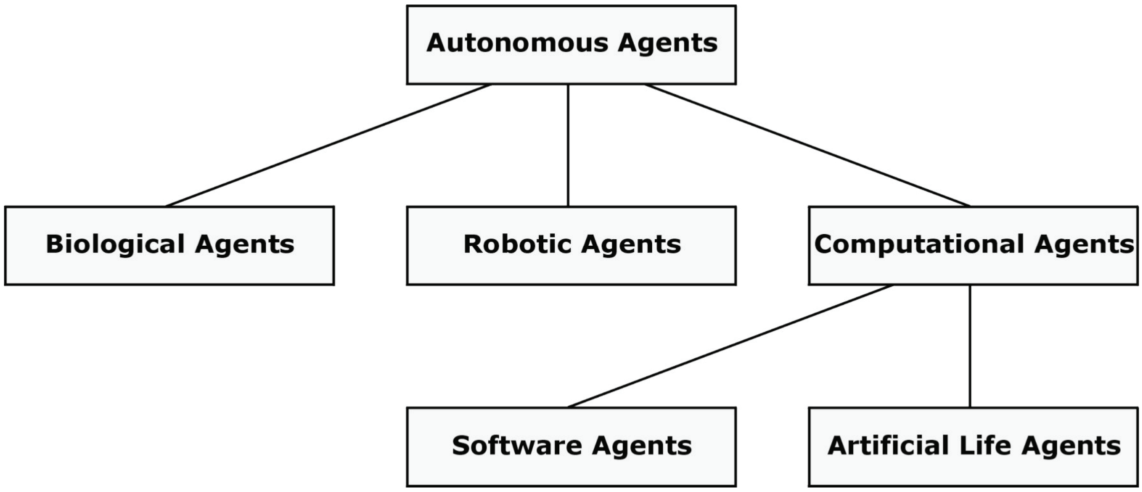

Figure 11. A natural kinds taxonomy of autonomous agents. Adapted from Franklin and Graesser (1996, Figure 1).

its previous experience, by consulting other agents, and from the user (Maes 1994: 33f.). If an agent can learn from the user, it is in a better position to adapt to the user's individual likes, dislikes, and procedures, which, in turn, allows it to become his or her *personalized* assistant (p. 32, cf. Section 3.2.5). Ways for agents to learn from the user include (pp. 33f.):

- *Observing and imitating the user*. The agent performs a continuous long-term observation of the user's actions to detect regularities and recurring patterns, and then suggests that it could automatically carry out these routine activities for the user.
- *Accommodating user feedback*. The user's response can be direct (the user explicitly communicates to the agent his or her opinion of its advice or actions) or indirect (the user follows or disregards a specific suggestion made by the agent), positive (e.g. "I would like to read more from this author") or negative (e.g. "This chart is too cluttered, clean it up").
- *Receiving explicit instructions (programming by example)*. Based on hypothetical examples of events and situations, the user instructs the agent what actions it should perform if such cases (or similar ones) occur. The examples may have different levels of specificity, ranging from descriptions of a single event or situation to templates (i.e. structures with slots for variable information) covering a whole class of events or situations. The agent integrates these examples and their associated actions into its knowledge base.

3.1.3 Believability

In the arts, a character is considered 'believable' if the audience can suspend their disbelief because the character successfully conveys the illusion of life (Thomas & Johnston 1984; Bates 1994: 122). Believable interactive characters with autonomy (cf. Section 3.1.1) that are situated in a computer-based environment are called *believable agents* (Bates 1994; Loyall 1997; Mateas 1997) or *synthetic characters* (Elliott & Brzezinski 1998).

Believable agents exhibit cues, behaviors, and properties that induce human observers and interaction partners to perceive (cf. Chapter 7.2) them as independent, sentient individuals rather than as cold, dependent, and uniform devices. Agents that seek to create and sustain the impression of being alive have to be able to act autonomously, proactively, responsively, and socially with respect to their environment in the first place. However, autonomy, proactiveness (cf. Section 3.1.6), responsiveness (cf. Section 3.1.7), and social ability (cf. Section 3.1.8) are also essential qualities of other types of intelligent agents which have to deal with a dynamic and unpredictable environment but are not necessarily believable. To achieve believability, an agent requires many more facets built on top of this foundation, which endow its persona with depth, appeal, and credibility:

- A strong, interesting *personality* (i.e. a set of psychological traits which defines the agent as a distinct individual, cf. Chapter 13.5);
- *Expressiveness* (i.e. the ability to convincingly and sensitively portray emotions (e.g. anger, happiness, surprise, etc.) and moods (e.g. cheerfulness, depression, optimism, etc.), cf. Chapter 13.4.2);
- *Empathy* (i.e. the ability to understand the state of mind or emotions of other individuals (to put oneself into their shoes) and to interact with them accordingly, either replicating their emotional state to communicate the ability to identify with them (*parallel empathy*) or expressing different emotions to influence their (possibly negative) state (*reactive empathy*) (McQuiggan et al. 2008);
- *Consistency* within and among appearance, behaviors, personality, etc. (cf. Chapter 7.5);
- *Idiosyncrasy* (i.e. the possession of distinctive and peculiar manners of speaking, holding and moving one's body, making gestures and facial expressions, etc., cf. Chapter 7.8);
- *Variability* in the kinds, timing, and manner of behavior (cf. Chapter 7.9);
- *Change* of attitudes, behaviors, moods, etc. over time;
- *Social relationships* with users and other agents (cf. Chapter 7.16).

The case for believable agents is usually made by arguing that an agent which users perceive as sentient rather than lifeless would provide a more agreeable communication and cooperation partner. Support for this view has been drawn from a series of studies which show that humans tend to respond socially to computers, like they would to other humans (Reeves & Nass 1998; Nass & Moon 2000, cf. Chapter 6.5).

Cassell and Tartaro noted that there are close ties between the believability of an agent and the agent's function or role (Cassell & Tartaro 2007: 406, cf. Chapter 7.14). Rather than striving to achieve the illusion of life in general, agent designers should be concerned with the more specific question "Is the agent we are building believable *as* a butler, salesperson, tutor, etc.?" (p. 406). That is, does the agent portray its appointed role in a way that makes people interact with it in ways similar to how they would deal with a human being in the same role? Cassell and Tartaro argued that realistic *behaviors* of an agent, especially in face-to-face communication (cf. Chapter 10.1 and Chapter 11), more so than its life-like appearance, may induce users to automatically suspend their disbelief and respond to the agent like they would behave toward a human interaction partner (pp. 406f., cf. Footnote 107).

3.1.4 Embodiment

Software agents working behind the scenes are typically bodiless entities, but there are other applications, including entertainment, education, and human-computer interaction, in which agents may show users a digital representation of themselves which is the result of carefully designing and integrating graphical and audio interface components to create an embodiment that is *anthropomorphic* (resembles human beings) in at least some respects. Typically, this representation is either a complete (i.e. fully articulate, with a face, hands, feet, and a torso, cf. Chapter 9.1) or partial virtual body (cf. Chapter 9.6) with a specific appearance (cf. Chapter

Table 3. The two major components of an embodied agent compared.

Agent (Program)	Character
• Back-end (functionality)	• Front-end (visible interface)
• Needs to be capable	• Needs to be pleasant
• Needs to have useful skills	• Needs to be a good communicator
• Needs to be adept, efficient, knowledgeable, or proficient	• Needs to evoke an emotional or intuitive response, such as to be believable, compelling, persuasive, or sincere
• Needs at least sufficient skills or knowledge	• No need to be particularly skilled or knowledgeable
• Performance glitches are not acceptable	• Behavioral glitches are not acceptable
• Must deliver good results	• Must deliver a good performance

9.2) and behavioral repertoire (cf. Chapter 10). The crafted embodiment defines the agent's identity and interactive potential.

When talking about such embodied agents, it is useful to draw a distinction between their background functional aspects (the *agent (program)*) and the embodiment through which they interact with the user (the *character* or *persona*). These two components of an embodied agent are compared in Table 3. Embodied agents also need to be separated from *avatars*, which represent and are controlled by human beings (rather than computer programs) in virtual environments (Bailenson & Blascovich 2004, cf. Figure 12).

Embodied agents are *interactive* if they can make use of their digital representation to act and react in relation to the user and the environment, which includes performing bodily actions to change the environment as well as verbal and non-verbal behaviors to communicate with the user. Embodied interactive agents are *conversational* if they can perform with their body like a human conversation partner in real-time multimodal dialogues face-to-face with human users (Bickmore & Cassell 2005). An *embodied conversational agent (ECA)* (Cassell et al. 2000a; Cassell 2001) uses its animated face and body for gaze, eye contact, facial expressions, gestures, locomotion, and body language, which support the agent's verbal messages (cf. Chapter 10.1). Given a suitably rich multimodal user interface,[15] the agent can both perform multimodal communicative actions and interpret such actions performed by the user in the context of the ongoing interaction. Embodied interactive agents are discussed in more detail in Section 3.3.

Embodied interactive software agents have been included in computer-based applications chiefly because they are expected to facilitate the interaction between the computer and the user (Dehn & van Mulken 2000: 2). In a recent survey of the role of interactive online characters in educational and commercial applications, Reeves (2002) argued that interactive characters can enrich human-media interactions, in particular online experiences, with social intelligence, taking advantage of users' natural social responses to interactive media (cf. Chapter 6.5). He identified the following benefits of interactive online characters (Reeves 2002):

[15] *Multimodal user interfaces* process combined user inputs from at least two modes (e.g. speech, face, gesture, writing, touch, etc.) and produce multimedia outputs that may be delivered by an embodied agent (Oviatt 2003: 287). Multimodal interfaces will be reviewed in detail in Chapter 11.2.

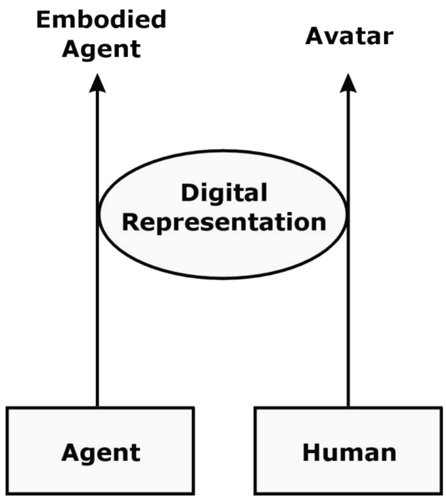

Figure 12. Digital representations in virtual environments may be controlled either by software agents or by human users. A representation puppeteered by a human being is called an 'avatar,' whereas embodied agents are completely program-controlled. The diagram shows this fundamental difference between avatars and embodied agents. Adapted from Bailenson and Blascovich (2004, Figure 1).

- *Characters make explicit the social responses that are inevitable* (Reeves & Nass 1998), by giving a social performance in interactions with the user that causes him or her to respond subconsciously in the same way.
- *Interactive characters are perceived as real social actors,* automatically and by default. As a result, characters can motivate, persuade, evoke sympathy, and so on (Fogg 2003).
- *Interactivity increases the perceived realism and effectiveness of characters.* Interactive characters can change their actions and responses, depending on user inputs and needs, in real time and as often as needed in the course of an interaction.
- *Interactive characters increase trust in information sources* through their social presence (cf. Chapter 5.4.5), by making contributions to the interaction that promote trust, by appearing and behaving in a way that is consistent with the task, and by enacting relevant social rules.
- *Characters have personalities that can represent brands.* Personality makes characters predictable and thus comprehensible (cf. Chapter 7.7) and agreeable partners in the interaction. It is possible to build characters with personalities that reflect corporate brands.
- *Characters can communicate social roles.*[16] Through appearance and behavior, they can show users who they are and how they will relate to them (as equals, subordinates, or superordinates) (cf. Chapter 7.14).
- *Characters can effectively express and regulate emotions.* They can use their emotional intelligence (cf. Chapter 13.4) to communicate emotions that are appropriate and helpful for the current user and task and thus contribute to a positive user experience.
- *Characters can effectively display important social manners.* They can make contributions to conversations with the user that are polite and culturally sensitive while also relevant, accurate, comprehensible, and sufficiently informative.

[16] A *social role* is a mutually recognized set of interconnected behaviors, rights, and responsibilities with respect to others, which are accepted for, expected of, and perceived in a particular individual in a given social status, position, and situation ([INT 36]).

- *Characters can make interfaces easier to use.* They provide an obvious and known place for assistance and information (Wexelblat & Maes 1997), and through their appearance and behavior they invite conversational interaction (Laurel 1990b: 359) that may combine spoken or written natural language with elements of non-verbal communication (cf. Chapter 10.1) to mimic features of face-to-face conversation, which humans are well equipped for and experienced in.
- *Characters are well liked* by most users, not just by particular groups. However, characters have to be carefully chosen and designed as some characters are liked more than others.

3.1.5 Mobility

Mobility refers to the ability of a software agent to travel from one environment to another (Lange & Oshima 1998: 50). Mobile agents (Nwana 1996; Lange & Oshima 1998; Kotz & Gray 1999; Lange & Oshima 1999; Eid et al. 2005) visit different places in a computer network to accomplish tasks on behalf of their users (e.g. buying or selling goods, managing computers and networks, doing research, etc.), which often involves collecting information from remote sources and interacting with remote services, and return with the results when the task has been completed (Nwana 1996, cf. Figure 13). Migration to another environment occurs when the program of a mobile agent decides to end its operation on the current computer and transports itself together with the results of its work to a destination computer of its choice in the network, where it carries on (Lange & Oshima 1998: 50).

Mobile agents have a number of advantages. For one, they can process remotely stored raw data in its location. As a result, only the results of processing have to be transferred over the network. Furthermore, mobile agents can take processing to another machine that has a smaller workload than the user's computer. In addition, they allow asynchronous computing. Users can launch mobile agents and 'forget' about them until the results come in (Nwana 1996). Further advantages of mobile agents are discussed by Nwana (1996) and Lange and Oshima (1998: 50f.). However, a number of issues need to be addressed in order to create successful mobile agents (Wayner 1995; Nwana 1996):

- *Transportation*. How is migration to another environment accomplished?
- *Authentication*. How can the identity of a mobile agent or its owner be verified?
- *Privacy*. How can the user make sure that a mobile agent is careful with his or her personal information and that no one else can read the information carried by the agent?
- *Security*. How can the user protect himself or herself against computer viruses brought back by his or her mobile agent? Or against 'hijacked' and 'turned' agents?
- *Payment*. How are agents going to pay for remote information and services? How can they be prevented from spending too much?
- *Performance*. How do large numbers of mobile agents at large affect network performance?
- *Interoperability*. How can mobile agents locate and access appropriate remote services and execution environments?

Mobility is not a defining characteristic of software agents. Neither does a piece of software require mobility to be called an agent nor does the possession of mobility alone justify the label 'agent' (Nwana 1996). In fact, mobile agents are classified as agents mainly because they have autonomy (cf. Section 3.1.1) and social ability (cf. Section 3.1.8) rather than mobility (Nwana 1996).

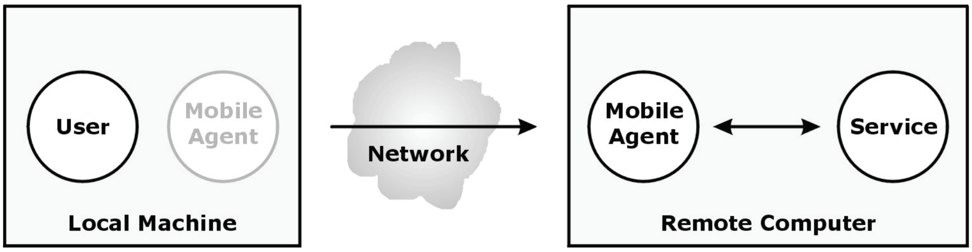

Figure 13. Mobile-agent operation. Adapted from White (1996). The mobile agent is dispatched by the user, traverses the network, and interacts with a service on a remote computer to accomplish its task.

3.1.6 Proactiveness

Rather than just responding to things happening in their environment, *proactive* software agents are capable of recognizing and seizing opportunities to take action in order to serve the user and their own agenda (Jennings et al. 1998: 8). The ability to take the initiative at appropriate points is an important aspect of autonomy (Nwana 1996, cf. Section 3.1.1). Proactive agents generate their own goals and attempt to achieve them through goal-directed behavior. They observe their environment to identify opportunities, and develop and execute plans of action to make use of them.

3.1.7 Responsiveness

In static environments, computer programs do not have to worry about their success or failure as the environmental conditions do not change, so they just execute blindly. In contrast, the environments of software agents are often dynamic and unpredictable, especially if they can be affected by humans and other agents. As a consequence, agents require the ability to perceive their environment and take appropriate action when it changes (Jennings et al. 1998: 8; Wooldridge 1999: 32). Responsiveness involves the following considerations:

- The agent has to determine which changes are relevant to its agenda.
- The agent may have to respond when only incomplete information is available.
- The agent has to deliver its responses in time; otherwise they may be useless.
- The agent must provide for failure.

3.1.8 Social Ability

In complex and dynamic environments, software agents often need the help of other agents to achieve their goals. Consequently, they require the ability to interact, when appropriate, with other artificial and human agents to complete their own tasks and to assist others with theirs (Jennings & Wooldridge 1996).

Social ability is the prerequisite for coordination and cooperation in *multi-agent systems*, i.e. networks of multiple agents working together to solve problems that a single agent cannot solve on its own (Sycara 1998: 80, cf. Chapter 14.1.5). The agents in a multi-agent system coordinate and cooperate by exchanging messages in a shared formal agent communication language (Wooldridge 2002: Chapter 8).

Interaction with the human user can involve (combinations of) different input and output modes: apart from typing on a keyboard or handling a mouse, people may use speech or gestures to interact with agents. In response, agents may appear as life-like characters on the screen, which convey internal states, messages, and other behavior through appropriate facial expressions or body language (cf. Section 3.1.4).

Social ability in the context of human-agent interaction also includes *emotional intelligence* (cf. Chapter 13.4). Agents with emotional intelligence can identify and interpret the emotions displayed by other (human or artificial) agents, communicate contextually and individually appropriate emotions, and manage emotions within others and themselves (Picard 2001).

So far, most interactions between users and intelligent software agents have been short encounters, but long-term relationships spanning multiple sessions over weeks and months are likely to become more common in the near future. Some examples of tasks that agents could carry out in short-term and long-term human-agent interactions are provided in Table 4. Social ability helps an agent to build a productive partnership with the user, both in the short run and in the long run. *Relationship-building skills* (cf. Chapter 7.16) are essential for agents that interact regularly with the same people over longer periods of time, such as interface agents (cf. Section 3.2.5) and pedagogical agents (cf. Chapter 4).

3.2 Types of Software Agents

As with their definition, there are multiple ways to classify software agents. One way is to use the set of properties discussed in the previous section. Furthermore, one can compare the complexity and expertise (cf. Chapter 12) of the agents. Software agents exhibit different levels of sophistication, which can be broadly defined as follows (from simplest to most complex) (Jennings & Wooldridge 1996; Jennings & Wooldridge 1998b: 18):

- *Gopher agents* are equipped with a set of pre-defined rules and assumptions which they apply to perform simple tasks. For example, the user could instruct an agent to notify him or her when the price of a certain product has dropped below a user-specified threshold, or to inform him or her about changes to a particular web site.
- *Service performing agents* receive well-defined high-level assignments from users, such as "Find me the cheapest plumbing service in London" or "Arrange a date for the final project review with the team members in the last week of November."
- *Predictive/proactive agents* are autonomous and flexible, which enables them to offer the user information or services unsolicitedly and opportunistically. For example, they can monitor the user's activities and suggest to take over recurring tasks. Or they have learned that the user likes a particular thing and inform him or her when they learn something interesting about that thing.

A second way to classify software agents is according to their purpose, function, or role (cf. Chapter 16). Software agents can play many types of roles in different applications (Jennings & Wooldridge 1998: 18). Some examples include the following (see Jennings and Wooldridge (1998b: 10–18) for a discussion of different applications of intelligent agents):

- *Buying/selling*. Agents purchase and trade goods and services on behalf of the user, for example participating as bidders or sellers in online auctions (cf. Section 3.2.2).
- *Decision support*. Agents collect information pertaining to a particular decision situation and prepare recommendations for human decision makers.

Table 4. Examples of agents' tasks in short-term and long-term interactions with users. Adapted from Park and Catrambone (2007, Table 1).

Short-Term Interaction	**Long-Term Interaction**
• Presenting information • Answering single questions • Recommending items or activities • Helping to carry out simple tasks • Giving guided tours	• Teaching the user complex knowledge or skills that require multiple sessions • Assisting the user throughout a long-term task or program • Becoming the user's companion

- *Digital butler/representative*. Agents take calls, respond to e-mails, receive visitors, etc. when the user is absent from the physical or virtual environment (Negroponte 1995: 149–152).
- *Entertainment*. Agents serve as allies or opponents in computer games or as interactive characters in movies or television (cf. Section 3.2.4).
- *Guides*. Agents show users around a new environment, such as a web site that the users visit for the first time, and introduce them to its features.
- *Information filtering and gathering*. Agents facilitate information management and reduce information overload by acquiring and filtering information according to the user's needs and capabilities (cf. Section 3.2.3).
- *Cognitive aid*. Agents aid or enhance human processes of remembering, thinking, and problem solving. For example, a calendar agent (cf. Section 3.2.5) manages the user's schedule and reminds him or her of important appointments or tasks. Pedagogical software agents can provide relevant advice, information, and scaffolding that facilitate learners' retention, transfer, and problem solving (cf. Chapter 12.3).
- *Network management*. Agents detect and handle congestions and intrusions, manage user privileges, etc.

Finally, software agents can be classified according to their underlying architecture. The following agent architectures will be discussed in more detail in Chapter 14.1:

- *Deliberative agents* generate alternative courses of action and choose one of them in order to achieve a goal. Deliberation involves manipulating complex representations of both the environment and courses of action (plans).
- *Reactive agents* do not maintain complex internal representations but map sensor input more or less directly to action output.
- *Hybrid agents* combine the strengths of the deliberative and the reactive approach in an architecture consisting of a hierarchy of interacting software layers.
- *Blackboard agent systems* comprise multiple knowledge sources that cooperate via a shared global workspace (the blackboard) to solve problems.
- *Multi-agent systems* are networks of cooperating agents that together solve problems which none of them could solve on its own.

The following sections review different categories or types of software agents, which have been established largely on the basis of the agents' properties and roles. The agents in each category may be of varying sophistication and may employ different architectures. A given agent may fit into more than one category; in that case, its actual classification would depend on what features or functions are emphasized.

3.2.1 Conversational Agents

Conversational agents make use of human language technologies[17] to interpret user inputs in natural language and to respond with natural language outputs, possibly combined with other input and output modes (Sidner 2002; Lester et al. 2004; McTear 2004). The architectures and component technologies of different types of conversational agents will be discussed in Chapter 11.3.

The purpose of conversational agents is to engage human users in spoken or text-based dialogues that are social or task-oriented in nature or mix both elements. Social dialogues (Bickmore & Cassell 2005) are conducted to build a relationship with users, to keep them occupied and interested while they are doing or waiting for something, and to discreetly pump them for more information about themselves. Task-oriented dialogues (Allen et al. 2001a) deal with various kinds of routine inquiries in areas such as the following (Lester et al. 2004: 10-2):

- *Customer service*. Agents answer routine questions about products and services.
- *Help desk*. Agents handle internal routine inquiries and problems of employees.
- *Web site navigation*. Agents show visitors to relevant sections of large web sites.
- *Guided selling*. Agents answer questions and offer guidance, especially to novice buyers of complex products.
- *Technical support*. Agents diagnose and solve technical problems interactively with users.

Conversational agents portrayed as animated anthropomorphic characters (cf. Chapter 9.1) can participate in multimodal face-to-face dialogues with users in which ideally both humans and agents can use coordinated verbal and non-verbal communicative behaviors (Bickmore & Cassell 2005). The non-verbal behaviors (gesture, face, gaze, posture, and locomotion) provide redundant or complementary information with respect to speech, and they have regulating functions in conversations (Bickmore & Cassell 2005, cf. Chapter 10.1). In social dialogues, non-verbal communicative behaviors contribute essential social cues, including attentiveness, positive affect, liking, and attraction (Bickmore & Cassell 2005).

3.2.2 E-Commerce Agents

Electronic commerce or *e-commerce* (Hoffmann et al. 1995; Chaffey 2006) uses computers and networks (nowadays in particular the Internet) as a platform which enables individuals and businesses to purchase, sell, market, and distribute products and services electronically. Either two businesses or a business and a consumer may be involved in an e-commerce transaction (*business-to-business, B2B* vs. *business-to-consumer, B2C*), during which they (automatically) exchange data, goods, payment, and services.

Intelligent software agents for e-commerce are (more or less) autonomous (cf. Section 3.1.1), adaptive (cf. Section 3.1.2), mobile (cf. Section 3.1.5), proactive (cf. Section 3.1.6), and socially competent (cf. Section 3.1.8). They can facilitate the following aspects of both B2C and B2B interactions (Maes et al. 1999; He & Leung 2002; He et al. 2003):

---

[17] *Human language technologies (HLT)* use knowledge about the structure of natural human languages and the world in tasks that require 'intelligent' processing of these languages, including dictation, document processing, human-computer interfaces, intelligent tutoring, and machine translation (Jurafsky & Martin 2008). HLT for conversational agents will be discussed in Chapter 11.1.

- *Need identification* (The buyer realizes the existence of an unfulfilled need). Agents can help users with routine purchases and make them aware of offerings that suit their preferences or habits.
- *Product brokering* (The buyer acquires information that helps him or her to choose a product). Agents can collect and filter product information given an identified need. Filtering techniques include *feature-based filtering* (matching product features against features on a wish list), *collaborative filtering* (recommending products bought by users with similar profiles), and *constraint-based filtering* (narrowing down the range of products based on user-specified constraints).
- *Merchant brokering* (The buyer finds the right seller for the product of interest). Agents can evaluate alternative vendors based on a comparison involving multiple attributes that may include price, delivery time, customer service, warranty, etc., in order to inform the user's decision of where to make the purchase.
- *Negotiation* (The buyer negotiates the price and other terms of the deal with the seller). Automated negotiations take place between agents representing their respective owners in order to work out contracts on the latters' behalf (Sandholm 1999: 84). The negotiating agents are *self-interested* (i.e. they choose actions with maximum utility for themselves) and operate under *bounded rationality* (i.e. their capacity for problem solving and information processing is limited) and *incomplete information* without centralized control (Jennings et al. 1998: 22).

An important consideration in doing business with agents and letting them do business on one's behalf is the building and maintaining of *trust* (Grandison & Sloman 2000; Patrick 2002). The New Oxford Dictionary of English defines trust as "the firm belief in the reliability, truth, ability, or strength of someone or something" (OED 1998: 1988). Trust is an interpersonal concept (Wexelblat & Maes 1997), "the condition in which one exhibits behavior that makes one vulnerable to someone else, not under one's control" (Zand 1972, quoted in Patrick 2002: 47). Trusting a software agent means that to a certain extent, users have to relinquish direct control and rely on the agent to perform its functions as expected when it carries out tasks on their behalf, thus rendering themselves vulnerable to any legal, monetary, or social risks associated with the agent's failure (Patrick 2002: 47).[18] The decision to trust an agent with buying, selling, or some other task inevitably involves balancing risk and trust: it has to be made by the user with regard to both the trust he or she puts in the agent's competence (and loyalty) and the costs of the agent's potential failure, which depend on the domain of interest.

[18] Users' feeling of being at the mercy of an autonomous computer program may become even stronger when they realize that unlike people, computers are not *moral agents*, which means that they are unable to take responsibility, make amends, or accept punishment for their actions, even if they harm people in some way (Fogg 2003: 218). The problem with software agents is that people have to trust entities that can act autonomously without being able to face the consequences if they do something unethical or illegal (p. 218). In addition, anthropomorphic embodied agents (cf. Section 3.1.4) could, in principle, be made believable (human-like) to a degree that would enable them to successfully mislead users both with respect to their own status and in commercial transactions (Heckman & Wobbrock 2000). If (and how) developers or owners can be held responsible and liable for the deeds of a software product with human-like appearance and behavior and/or capable of making decisions by itself is still an unresolved issue.

Trust is not an instant relationship but needs to be given time to develop between users and the agents carrying out tasks on their behalf (Jennings & Wooldridge 1998b: 10). A given agent can facilitate the process of trust building by showing:

- *Believable* behaviors (cf. Section 3.1.3), in particular rituals involved in conversation, such as greetings, small talk, and farewells (Cassell & Bickmore 2000);
- *Predictable* behaviors (Lewis 1998: 73f.; Heckman & Wobbrock 2000):
  - The agent exhibits consistency of behavior in response to input (cf. Chapter 7.5).
  - The agent's behavior is comprehensible in terms of its underlying causes and rules (cf. Chapter 7.7).
  - The user understands the agent's function, competencies, and dependability.

While trying to win the user's trust, the agent has to walk a tight-rope between seeking too much and too little guidance from the user, i.e. between needlessly annoying the user and over-confidently overstepping its bounds (Jennings & Wooldridge 1998b: 10). In general, agents have to know both their capabilities and limitations, as well as communicate them to the user (Jennings & Wooldridge 1998b: 10; Heckman & Wobbrock 2000).

3.2.3 Information Agents

An ever-increasing volume and diversity of information can be stored, distributed, and linked using today's computers and computer networks (in particular the Internet and its most widely used service, the World Wide Web). Given their limited resources, people cannot keep up with this development, which results in the experience of feeling lost (not finding anything) in the vast network of information or of being overwhelmed by an excess of information (*information overload*;[19] Edmunds & Morris 2000; Eppler & Mengis 2004). Information agents are designed to help humans to process information by facilitating information management and reducing information overload (Nwana 1996). They are equipped with the ability to access distributed and heterogeneous information sources (e.g. legacy databases, the World Wide Web, or other information agents) and to acquire, collate, and manipulate information on behalf of the user or other agents, preferably just in time (Jennings & Wooldridge 1996; Nwana 1996; Sycara et al. 1996, cf. Figure 14). Information agents possess attributes including autonomy, adaptiveness, mobility, proactiveness, and social ability to varying degrees (Nwana 1996).

Information agents maintain representations of available information sources. In addition, they possess strategies to select and access these sources, handle conflicts and inconsistencies, and integrate different pieces of information (Nwana 1996; Sycara et al. 1996). An information agent takes action either in response to *queries* formulated by users or other agents or when it detects certain patterns while *monitoring* a set of information sources (Sycara et al. 1996). Information agents can perform both *information filtering*, allowing only relevant information to reach the user, and *information gathering*, proactively searching information sources and fetching relevant information on the user's behalf (Jennings & Wooldridge 1996; Sycara et al. 1996; Jennings et al. 1998).

Much information is still stored as natural language text in documents, which is more accessible to humans than to machines. As a consequence, information agents require effective

[19] The term first appeared in a book by the sociologist and futurologist Alvin Toffler (1970).

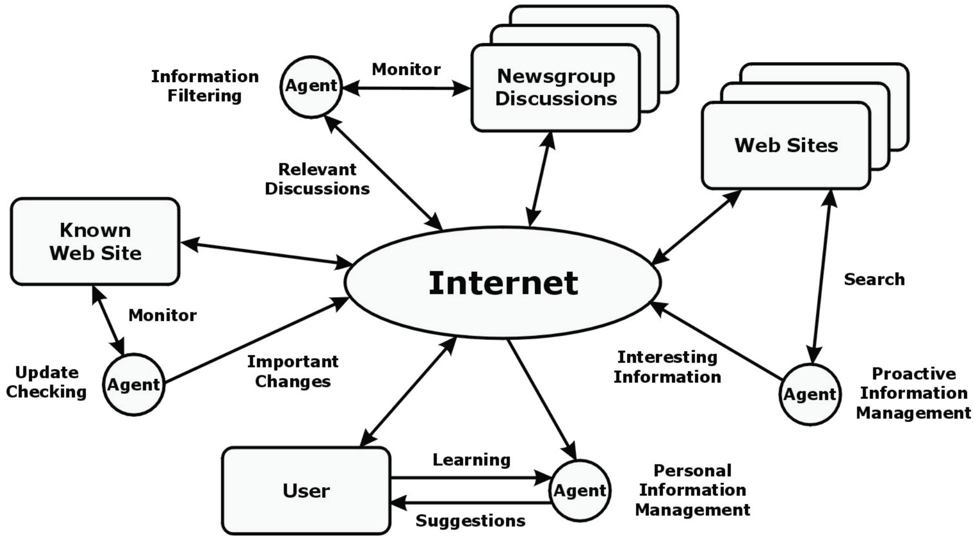

Figure 14. Agent-based information management on the Internet. Adapted from Jennings and Wooldridge (1996, Figure 1).

ways to process the *content* (rather than the form) of text documents. These content-based document processing techniques include the following:

- *Information retrieval* finds and retrieves documents to satisfy user queries for information (Baeza-Yates & Ribeiro-Neto 1999; Tzoukermann et al. 2003; Grossman & Frieder 2004).
- *Information extraction* identifies and extracts selected types of entities, relations, or events in documents (Cowie & Wilks 2000; Grishman 2003; Cunningham 2006; Sarawagi 2008).
- *Question answering* finds answers to users' questions in a large collection of online documents (Harabagiu & Moldovan 2003; Maybury 2004).
- *Automatic summarization* condenses the contents of one or more documents in such a way that the resulting document is much shorter while preserving the important information of the original document(s) (Mani 2001; Hovy 2003; Das & Martins 2007).

In e-learning (cf. Chapter 2), information agents can serve as an adaptive and interactive research tool that provides a uniform interface to the various resources (text, graphics, animation, audio and video elements) of a web-based learning environment. These *virtual librarians* (Stephenson 1993: 99–103) use both local resources and others they find in external sources, such as the World Wide Web, to retrieve relevant documents, extract information of interest, provide concise responses to learners' questions, and create summaries of extensive resources, all in response to content-based (rather than keyword-based) queries for information.

3.2.4 Game Agents

For a long time, development of computer games has focused on improving the quality of game graphics. However, with computer graphics approaching the level of photorealism, less and less innovation can be expected from that aspect of computer games. Therefore, in recent years, voices have multiplied which argue that *artificial intelligence (AI)*[20] will be an essential ingredient of quality computer games in the future (Laird & van Lent 2000; Laird 2001; Laird & van Lent 2001; Isla & Blumberg 2002; Mateas 2003; Narayek 2004).

The primary role of AI in computer games is seen in providing intelligence for *non-player characters (NPCs)*, including friends, foes, guides, mentors, traders, and other entities,[21] which have different goals, knowledge, and capabilities (Laird & van Lent 2001: 16; Narayek 2004: 60). Humanoid NPCs should be endowed with similar strengths (and limitations) as the human player, in terms of accuracy, creativity, flexibility, perception, and skill (Laird & van Lent 2001: 22). They should also be virtual individuals that players enjoy interacting with, so they need an appropriate physical appearance, communicative capabilities, social graces, and so on (Narayek 2004: 60). In other words, the goal of designing entities with artificial intelligence for games is not necessarily to produce a 'perfect' NPC but one that is as believable and enjoyable for players as possible (Narayek 2004: 60). Detailed requirements for different AI roles in computer games are shown in Table 5.

3.2.5 Interface Agents

Interface agents (Laurel 1990b; Maes 1994; Negroponte 1995: Chapter 12; Shneiderman & Maes 1997; Wexelblat & Maes 1997; Dehn & van Mulken 2000; Middleton 2001) are digital personal assistants designed to facilitate human-computer interaction by cooperating with the user to accomplish tasks in a computer-based environment (Nwana 1996). Unlike other kinds of intelligent user interfaces, interface agents are autonomous (cf. Section 3.1.1), adaptive (cf. Section 3.1.2), proactive (cf. Section 3.1.6) (Nwana 1996), and have social ability (cf. Section 3.1.8). They determine users' habits, interests, and preferences and are authorized to act on their behalf to perform (often mundane, tedious, or repetitive) tasks that have been delegated to them (e.g. scheduling meetings, filtering e-mails, booking flights, etc.) (Maes 1994; Nwana 1996).

Interface agents thus change the nature of users' interaction with computers from direct manipulation to indirect management (Kay 1990) and from user control to cooperation of users and agents in which both monitor the environment, engage in activities, and take the initiative in communication (Maes 1994: 31). This paradigm shift in human-computer interaction is hypothesized to suit novice and non-expert users in particular because they can benefit from the proactive assistance provided by software agents to help them to accomplish large and/or complex tasks (Nwana 1996). In addition, embodied interface agents with the ability to perform (and interpret) verbal and non-verbal communicative behaviors are expected to increase the naturalness of human-agent interaction (cf. Section 3.1.4).

This optimistic view of the potential of interface agents is not shared by everyone, though. In fact, interface agents have a number of outspoken critics, who argue that humanizing the interface through (embodied) agent assistants might actually impair, rather than benefit, human-computer interaction because (Shneiderman 1995; Lanier 1995; Erickson 1997; Norman 1997; Shneiderman 1997; Shneiderman & Maes 1997):

---

[20] *Artificial intelligence (AI)* aims to give computers and other devices the ability to perform tasks that people accomplish by applying their intelligence, such as chess playing, tutoring, medical diagnosis, natural language understanding, and face recognition (Cawsey 1998; Russell & Norvig 2003; Luger 2005). The acquisition, representation, and use of *knowledge* has been a major concern in the field of AI throughout its history (cf. Chapter 12.1).

[21] Common social roles of non-player game characters are discussed by Isbister (2006: 229–250).

Table 5. Roles of non-player characters with artificial intelligence and their requirements. Compiled from Laird and van Lent (2001: 21–24).

Role	Requirements
Tactical Enemies	• Autonomy • Reactive behavior, integrated planning, and commonsense reasoning for interacting with complex and dynamic environments • Models of high-level vision with human-like abilities and limitations • Navigation, including path planning, spatial and temporal reasoning • Ability to plan, counter-plan, and adapt to enemies' strategies and tactics • Movement and reaction times similar to humans
Partners	• Cooperation and coordination with the human player, possibly using (spoken) natural language and gestures • Understanding of teamwork • Modeling the goals and adapting to the style of the human partner
Support Characters	• Portrayal of a supporting human role in the game environment • Ability to interact with and adapt to the environment, human players, and other support characters • Human-like realistic movement, personality, emotions, natural language understanding and generation
Strategic Opponents	• Low predictability • Ability to generate a cohesive high-level strategy (requires integrated planning, commonsense reasoning, spatial reasoning, plan recognition, and counter-planning) • Resource allocation (requires scheduling and temporal reasoning) • Playing with realistic reaction times and movements
Units (teams, armies, etc.)	• (Semi-) Autonomy • Path planning and path following • Commonsense reasoning • Coordination with other units
Commentators	• Ability to describe the ongoing game event in natural language • Plan recognition to uncover the strategy and tactics of teams and players • Deep knowledge of the game

- It causes users to form incorrect mental models of an agent (cf. Chapter 6.3). As a result, by way of overgeneralization from existing human-like qualities of the agent, users may come to think that the agent possesses attributes and capabilities which it actually does not have (Norman 1997).
- Many tasks in human-computer interaction (e.g. writing a letter or preparing a spreadsheet) can be accomplished more efficiently without social interaction with interface agents (cf. Section 3.1.8) because social interaction is not necessary for these tasks (Prendinger & Ishizuka 2004b: 6, cf. Chapter 8.3).
- People like feeling in control of the artifacts with which they interact (Shneiderman 1995; Norman 1997; Shneiderman 1997). This preference is not immediately compatible with the

need to relinquish control to an autonomous agent and to trust the agent that it will perform the task to the user's satisfaction.[22]

- Users have a preference for predictable systems (Shneiderman 1997). However, agents have some properties (in particular autonomy and proactiveness, cf. Section 3.1.1 and Section 3.1.6) that can make their behavior unpredictable.
- Adding an (eye-catching) object to an interface, in particular if it keeps unduly drawing attention to itself, e.g. by performing behaviors that are not related to the current task, may introduce a source of distraction for users (Walker et al. 1994), increasing their extraneous cognitive load and consequently impairing their performance (Paas et al. 2003; Sweller 2005a; Clark et al. 2006, cf. Chapter 2.2.2).
- Humanizing the interface can lead to dangerous dependency relationships with machines. Lanier argued that the use of intelligent software agents would cause people to "redefine themselves into lesser beings," by making themselves 'dumb' in order to make the agent appear 'smart' (Lanier 1995: 67f.).

These are valid concerns which need to be considered in the design of interface agents. They imply that interface agents should make their capabilities and limitations transparent to the user (cf. Section 3.2.2); engage in social interaction with the user only when the task requires it; make sure that they do not take control of the interaction and the task from users against their will (Wexelblat & Maes 1997); win the user's trust by performing predictably and reliably (cf. Section 3.2.2); refrain from distracting the user with idle, irrelevant behaviors (cf. Chapter 10.2); and adopt a subservient rather than dominant role in the relationship with the user. Table 6 lists challenges that have to be addressed in order to achieve successful cooperation between agents and users.

The architecture of an interface agent is shown in Figure 15. Interface agents interact with their employer (the user), applications (on behalf of the user), and other agents. As indicated above, adaptiveness is an important capability of interface agents, which helps them to improve their services for particular users over time. The different options for agents to learn that are illustrated in Figure 15 were already discussed in Section 3.1.2. To interact with users, interface agents can use different input and output modes, alone or in combination. When they deal with other agents, formal agent communication languages are used (cf. Section 3.1.8). Internal states and messages can be conveyed by a life-like animated character. For example, Peedy the Parrot (Ball et al. 1997) is an embodied interface agent that combines spoken language understanding (cf. Section 11.3.3.1), dialogue management (cf. Chapter 11.3.3.2), 3D animation (cf. Chapter 9.4), synthetic speech output (cf. Chapter 10.1.1 and Chapter 11.1.2), and sound effects in a conversational assistant with human-like attributes and behaviors (while portrayed as a green parrot), which can interact with the user in task-oriented spoken dialogues (cf. Section 3.2.1).

---

[22] Schiaffino and Amandi found that users are distrustful of completely autonomous interface agents. The vast majority of users prefer to remain in control of their computer system and would rather not have an agent autonomously execute actions without their knowledge and consent, although they do not seem to be adverse to the idea of having a personal agent assistant in principle (Schiaffino & Amandi 2004: 135ff.).

Table 6. Challenges involved in the interaction between users and interface agents (Middleton 2001).

Knowing the user	• Identifying user goals through observation and feedback • Placing user goals in context • Adapting to changing user goals • Minimizing initial training time and effort
Interacting with the user	• Deciding on the agent's degree of control over tasks • Developing the user's trust in the agent • Finding a suitable metaphor for agent-based interfaces • Making agent-based interfaces easy to use for novices
Competence in helping the user	• Timing interruptions of the user appropriately[23] • Accomplishing tasks autonomously to the user's satisfaction • Determining how to (partially) automate tasks

3.3 Embodied Interactive Software Agents

Embodied interactive software agents are complex software systems equipped with a digital embodiment (cf. Section 3.1.4) and the ability to make use of this embodiment to communicate verbally and non-verbally with the user and to perform bodily actions in the environment in which they are embedded. To accomplish their tasks, these agents interact with users, other agents, objects, and events in their environment as well as with software applications and database systems that provide necessary information and services. Interaction involves sensing the environment and acting on it (which includes the performance and interpretation of bodily actions) to communicate with the user or with other agents or to proactively participate in the environment, both in pursuit of a user- or a self-assigned agenda.

The general architecture of an embodied interactive agent is shown in Figure 16. It consists of a pipeline of five processing steps that accepts communicative events and other sensory percepts as input and produces communicative and other actions as output. The peripheral steps deal with the recognition of sensor inputs and the synthesis of action outputs, respectively, while the internal stages are concerned with interpretation, internal processing, and behavior generation. The architecture also enables the agent to bypass the internal processing stages in

[23] McFarlane (1999) argued that users are sensitive to interruptions from computer-based systems, including intelligent software agents, due to their limited cognitive abilities. He identified four ways to time interruptions: (1) *immediate* (the agent interrupts the user whenever it sees the need, requiring him or her to attend to the matter at once); (2) *negotiated* (the agent lets the user know that his or her attention is needed and negotiates the time of dealing with the interruption with the user); (3) *mediated* (the agent informs a mediator about its need to interact with the user, and the mediator decides when and how the agent may interrupt the user); (4) *scheduled* (the agent's interruptions are timed according to a pre-defined schedule, for example once per hour) (McFarlane 1999). Bickmore et al. (2008) found evidence for the superiority of a fifth method, *empathic* interruptions, in which the agent apologizes for the interruption, asks about the user's emotional state, and expresses empathy. In general, knowing when and how to interrupt the user is critical for agents' success: a high level of perceived intrusiveness of an agent has been shown to lead to user frustration and the agent's early unemployment (Serenko 2006).

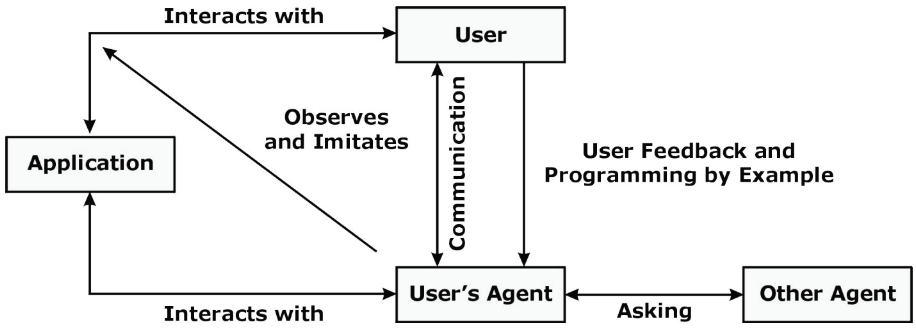

Figure 15. The architecture of an interface agent. Adapted from Nwana (1996, Figure 4).

order to produce immediate (hardwired) reactions to environmental stimuli (Cassell et al. 2000b: 43). The processing steps draw on several knowledge sources and maintain different kinds of models. The elements of the architecture (processing stages, knowledge sources, models, and back-end systems) are briefly described below:

Processing Stages

- The *sensory module* integrates technologies that enable the agent to perceive its environment and to accept user inputs. Examples include technologies for speech recognition; face and gesture recognition and tracking; monitoring objects, events, movement, etc.; and handling keyboard and mouse actions.
- The *interpreter* fuses (combines) all percepts into a coherent and consistent interpretation of the incoming communication, events, and information about the state of the environment (cf. Chapter 11.2).
- *Internal processing* includes dialogue management, decision making, interaction with external applications and databases, etc.
- *Behavior generation* assembles action sequences consisting of speech, gesture, movement, etc. that, if carried out, achieve communicative or task-related goals.
- The *character player* is concerned with the rendering of animations and the synthesis of speech. In particular, it delivers the animations and verbal utterances of the agent persona.
- *Hardwired reactions* include instant behavioral responses of the agent to incoming stimuli, which are produced without passing the input through the internal processing stages (Cassell et al. 2000b: 43).

Knowledge Sources

- *Pattern databases* for speech; faces and facial expressions; gaze patterns; different kinds of gestures, body postures, and locomotion; object shapes; etc. are required for the low-level processing (detection, tracking, and recognition) of these different types of input.
- *Domain knowledge* is knowledge about the agent's area of expertise.
- *Linguistic knowledge* comprises knowledge about language understanding and generation (grammars, lexicons, etc.).
- *Communicative knowledge* includes knowledge about how to participate effectively in face-to-face conversations, including initiation and termination, turn-taking, and feedback.
- *Affective knowledge* enables the agent to interpret, express, and communicate or simulate emotions, moods, personality, etc.

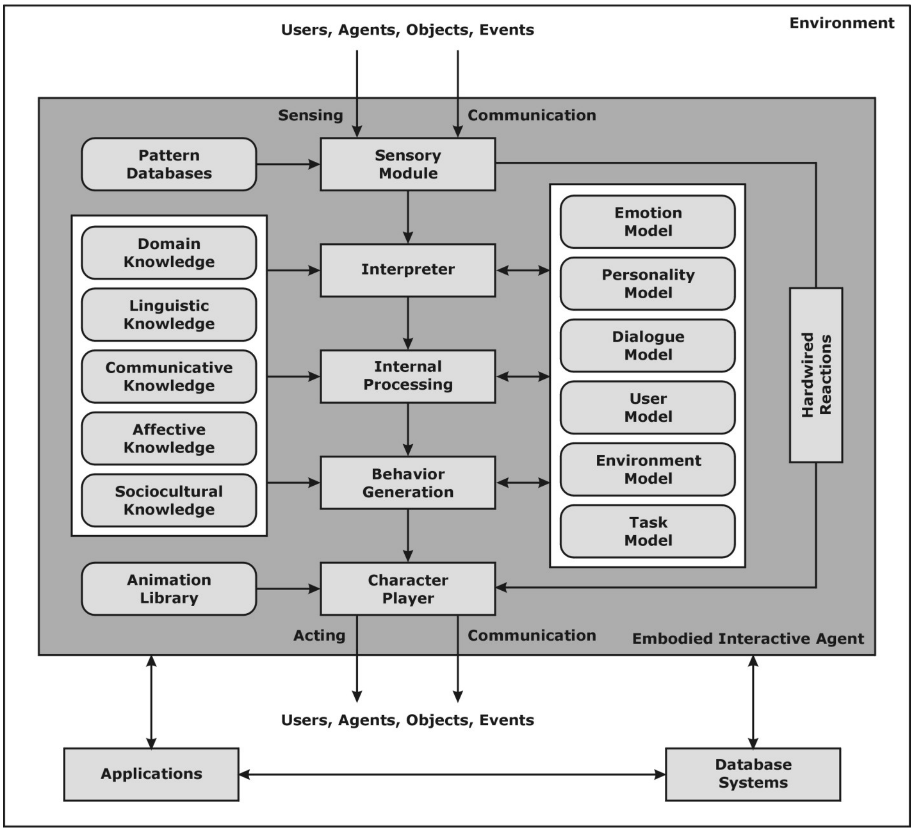

Figure 16. The architecture of an embodied interactive software agent.

- *Sociocultural knowledge* includes aspects of the target culture, such as traditions, gender roles, display rules, etc.
- An *animation library* provides data and/or procedures at varying levels of complexity for creating animations.

Models

- The agent's *emotion model* is a representation of its own simulated or communicated emotional state, i.e. how the agent 'feels.'
- The *personality model* represents the combination of long-term psychological traits that distinguishes the agent from other individuals but also makes it predictable to some extent.
- The *dialogue model* is a continuously updated model of the ongoing conversation between the agent and the user (or another agent).
- The *user model* captures the agent's evolving view of various characteristics of the user (e.g. goals, knowledge, and preferences).
- The *environment model* is a continuously updated representation of users, agents, objects, and events in the environment.
- The *task model* is a continuously updated record of the current task, including its history, current state, and next steps.

Back-End Systems

- *Applications* include software systems that provide specific functionality and services which help the agent to carry out its tasks.
- *Database systems* provide facilities for storing, accessing, and manipulating large quantities of data in an efficient and organized way.

The architecture in Figure 16 presents a big-picture view of the components that have to be developed for embodied interactive software agents. How they can be effectively designed for embodied interactive pedagogical agents will be discussed in detail in Chapter 9 to Chapter 14.

3.4 Summary

This chapter introduced the concept of software agents, discussed their properties, and reviewed different types of software agents. Intelligent software agents were defined as software systems that are situated in a computer-based environment and have the capacity for autonomous, flexible, and goal-directed interaction with this environment. They can detect relevant states of their environment and affect it through their actions. Autonomous software agents follow their own agendas to achieve desired results, for themselves or in the service of a human user. The following properties of intelligent software agents were discussed in Section 3.1:

- *Autonomy*. The ability to continually act independently of direct user control;
- *Adaptiveness*. The ability to acquire, change, and improve behaviors;
- *Believability*. The ability to create the illusion of life;
- *Embodiment*. The ability to interact through a digital representation;
- *Mobility*. The ability to migrate from one environment to another;
- *Proactiveness*. The ability to recognize and seize opportunities for action;
- *Responsiveness*. The ability to respond to changes perceived in the environment;
- *Social ability*. The ability to interact with human users and other software agents.

Software agents have been developed for different applications and roles. In Section 3.2, several types of software agents were described: conversational agents (cf. Section 3.2.1), e-commerce agents (cf. Section 3.2.2), information agents (cf. Section 3.2.3), game agents (cf. Section 3.2.4), and interface agents (cf. Section 3.2.5).

The third section of the chapter discussed embodied interactive software agents in some detail, characterizing them as complex entities that interact (through bodily actions) with users, other agents, objects, events, software applications, and database systems in their environment in order to accomplish their tasks. A general architecture of an embodied interactive agent was presented in Figure 16. The components of this architecture, including processing stages, knowledge sources, models, and back-end systems, were also described in Section 3.3.

4 Pedagogical Agents

Pedagogical agents are embodied[24] interactive software agents embedded in interactive learning environments that work with students to facilitate their learning. These agents can promote effective and efficient learning by providing the individual learner, groups of learners, and/or other agents with contextualized, qualified, personalized, and timely assistance, cooperation, instruction, motivation, and services. Pedagogical agents are based on previous research and development in the fields of intelligent tutoring systems (Johnson 1998, cf. Chapter 2.3.3), software agents (cf. Chapter 3), and animated characters (Thomas & Johnston 1984; Bates 1992; Elliott & Brzezinksi 1998; Prendinger & Ishizuka 2004a; Isbister 2006).

Appropriately designed pedagogical agents are equally suitable for supporting the learning of both abstract subject matters (e.g. foreign languages, history, or physics) and physical, procedural tasks, like those that have to be performed to operate and maintain complex machinery (Rickel & Johnson 2000; Johnson et al. 2000; Lester et al. 2001). Figure 17 shows representative examples of animated pedagogical agents that have been developed in recent years. More information about these agents can be found in the references cited. While still a fairly young field of research, a number of surveys of pedagogical agents have already been published (Johnson 1998; Johnson et al. 2000; Johnson et al. 2001; Person & Graesser 2003a).

The first section of this chapter[25] provides an overview of the capabilities of pedagogical agents and shows how they can use these capabilities to assist human learners. The second section is concerned with the requirements which pedagogical agents with the previously discussed capabilities have to meet. The chapter concludes with a summary in Section 4.3.

4.1 Capabilities of Pedagogical Agents

Pedagogical agents enrich the learning experience with their capacity to act (Lachman 1997: 32), to perform a wide range of instructional interactions (cf. Chapter 12.3) in order to promote learning (Johnson 1998). They add elements of human face-to-face instructional conversations to computer-based learning environments, helping learners with their knowledge of the subject matter, instructional strategies, and available resources for learning, and conveying enthusiasm for the subject matter and empathy for the learner by communicating thoughts and emotions (Johnson 2003: 78).[26] Personifying the interface to the learning environment, they exploit the natural human tendency to respond socially to computers (Reeves & Nass 1998; Shaw et al. 2004, cf. Chapter 6.5).[27] Furthermore, Prendinger and Ishizuka argued that character-based

---

[24] There are, of course, pedagogical agent roles which do not require an embodiment. An example would be an agent in an administrative role, which usually works behind the scenes and does not have to interact ‘face-to-face’ with learners (cf. Chapter 12.3.14).

[25] Substantial parts of the following discussion were previously published in Franke (2006).

[26] Kim et al. reported that pedagogical agents as learning companions (PALs) which showed empathy in response to learners’ emotional state led the latter to demonstrate more interest and self-efficacy in tasks than learners working with a PAL that did not respond to their expressed emotions (Kim 2005; Kim et al. 2007).

[27] Nass and Sundar (1994) listed four cues which induce automatic social responses in people: (a) use of language; (b) interactivity; (c) playing roles traditionally filled by people; (d) voice. These cues will be discussed in subsequent chapters of this thesis.

interfaces, including pedagogical agents, can take advantage of users' previous knowledge about how to participate in face-to-face interactions with other humans:

> *[T]he power of character-based [interfaces] derives from the fact that people know how to interact with other people by using the modalities of their body (voice, gesture, gaze, etc.) and interpret the bodily signals of their interlocutors. Hence, character-based interfaces aim at realizing embodied interaction and intelligence (...) rather than interaction with 'invisible' devices ... (Prendinger & Ishizuka 2004b: 6).*

Human instructors naturally make use of their bodies to support their teaching activities. Learners, in turn, benefit from the rich sensory experience of interacting with the instructor (and also with their fellow learners) through the means and mechanisms provided by their bodies, including gestures, facial expressions, voice quality, gaze, posture, and locomotion (cf. Chapter 10.1) apart from the actual instructional messages and behaviors. A significant drawback of current online courses (cf. Chapter 2.4.2.2) is the participants' lack of access to bodily features and behaviors of other participants (Woodill 2004). Embodied pedagogical agents can be designed to address this deficiency by engaging in bodily (inter-) actions in a way that benefits information and communication processes in learning environments (Gulz 2004).

Pedagogical agents combine capabilities that enable them to serve important instructional, communicative, and motivational functions for learners:

- *Monitoring*. Embedded in learning environments, pedagogical agents can monitor learners' actions and intervene with appropriate, contextualized problem-solving assistance when needed (Lester et al. 1997b).
- *Flexibility*. A pedagogical agent has the ability to tune its behavior to changes in the learner and the learning environment (Johnson 1998). It evaluates the learner's progress and seizes opportunities for instruction which occur in the course of the learner's activities (Johnson 1998; Johnson et al. 2000).
- *Explanation*. Pedagogical agents can give each learner contextualized and customized explanations of concepts and procedures, on demand or when they perceive that the learner requires assistance (Johnson et al. 2000).
- *Demonstration*. People often find learning a new task easier when they can watch an expert do it first (Isbister 2006: 146). According to social learning theory (Bandura 1977), human learning is facilitated in social contexts that allow people to watch and learn from each other, including their errors and achievements (Isbister 2006: 148). Embodied agents have the ability to demonstrate to learners how to solve problems and perform tasks by carrying out physical actions in the environment, which may be combined with explanations of what the agent is doing, and why (Johnson 1998; Johnson et al. 2000).
- *Dialogue*. Embodied interactive pedagogical agents (cf. Chapter 3.3) can engage learners in a multimodal dialogue (Johnson 1998) that involves the coordinated use of interdependent verbal and non-verbal communicative behaviors (speech, facial displays, gaze, gestures, postures, and locomotion, cf. Chapter 10.1) by both parties. This emulation of human face-to-face interaction provides a digital equivalent to the socially interactive aspects of face-to-face learning environments (Dowling 2000; Gulz 2004).
- *Individual and team support*. Pedagogical agents can serve both individual learners and teams of learners, by tutoring team members on their roles and filling in for missing team members (Rickel & Johnson 2003).

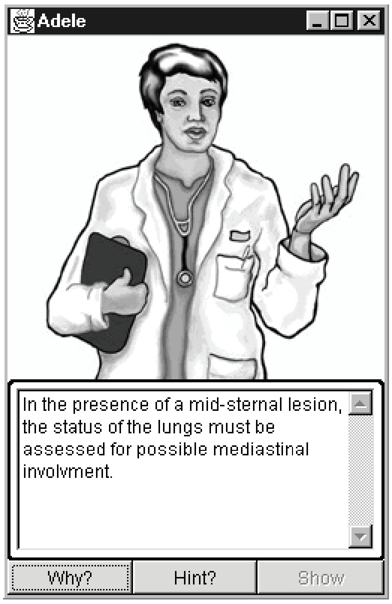

Figure 17a. Adele is a web-based pedagogical agent that acts as a tutor and guide for learners engaged in problem solving in the fields of medicine and dentistry (Shaw et al. 1999; Johnson et al. 2003a). Screenshot from Johnson et al. (2000: 50, Figure 2).

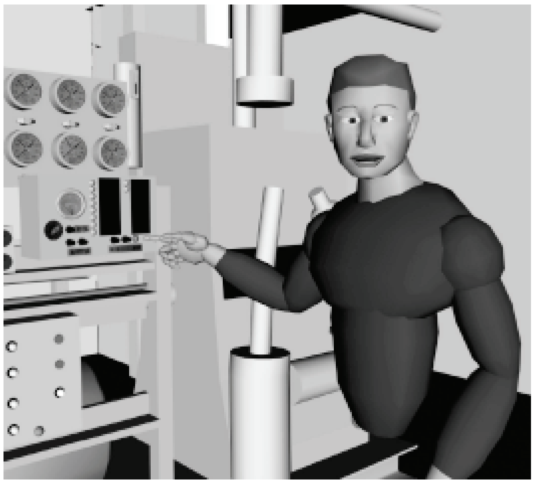

Figure 17b. Steve is an animated tutor agent that interacts with students in a virtual reality training environment to help them to learn how to perform operation and maintenance tasks (Rickel & Johnson 2000). Screenshot from Johnson et al. (2000: 49, Figure 1).

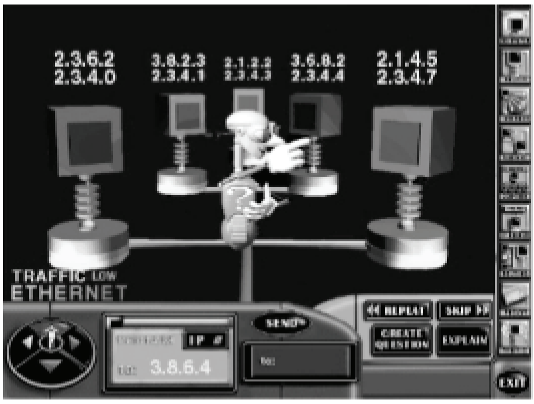

Figure 17c. Inhabiting a virtual environment of networks and routers, Cosmo plays the role of a coach advising learners in their decisions about how to route packages of data through the simulated network. Cosmo explains both subject matter and problem-solving actions to the learner (Lester et al. 2001). Screenshot from Johnson et al. (2000: 51, Figure 4).

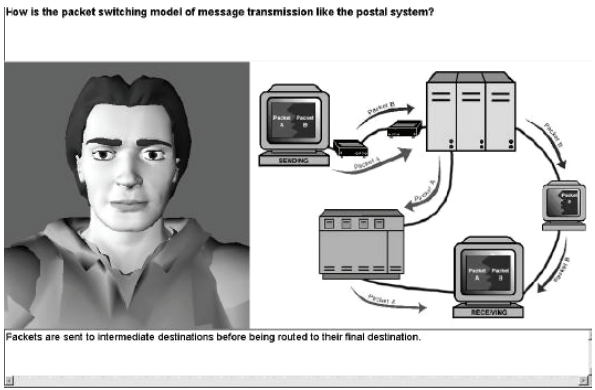

Figure 17d. AutoTutor is an intelligent web-based tutor that provides an animated conversation partner which engages the learner in a tutorial dialogue. Versions of AutoTutor have been developed for the domains of computer literacy and physics (Graesser et al. 2001; Graesser et al. 2005). Screenshot from Graesser et al. (2001: 40, Figure 1).

- *Motivation.* Pedagogical agents with emotional intelligence (cf. Chapter 13.4) can encourage learners by demonstrating a genuine interest in them and their progress, take action if they sense learner frustration, show enthusiasm for the content, and increase learners' enjoyment (Elliott et al. 1999; Gulz 2004).

4.2 Requirements of Pedagogical Agents

Since pedagogical agents are software agents, they must satisfy many requirements of the latter (Johnson 1998). They have to be able to perceive their environment in order to detect the need for intervention or opportunities for learning. Furthermore, they have to decide intelligently and independently on the actions to take either proactively (e.g. to prevent the learner from making a mistake) or in response to a multitude of environmental stimuli (e.g. a learner question). Their behavior has to be coordinated with the behavior of other (human and computer) agents (Johnson et al. 2000), which is particularly important for cooperative and team learning (Rickel & Johnson 2003). In addition, pedagogical agents must be able to adapt and improve over time through learning (cf. Chapter 3.1.2). Pedagogical agents also require flexibility in processing and behavior to deal appropriately with unforeseen situations and learner actions.

Furthermore, the instructional role of a pedagogical agent may involve the need to engage in face-to-face interactions (Cassell 2000a) with learners (and possibly other virtual actors, cf. Chapter 16). This is only possible if the agent has a body with the necessary communicative capabilities. Pedagogical agents designed as embodied interactive software agents have to integrate graphical and audio interface components into a digital representation (cf. Chapter 9) that possesses at least some aspects of the appearance and behavioral repertoire of a human body and enables the agent to perform bodily actions in its dealings with the learner and the environment (cf. Chapter 3.1.4). Embodied conversational pedagogical agents should be able to participate in instructional dialogues with the learner in which they interpret and produce coordinated verbal and non-verbal communicative behaviors (cf. Chapter 10.1 and Chapter 11).

Over and above the general and communicative requirements on pedagogical agents, their mission to facilitate human learning necessitates further characteristics which may not be found in other agents:

- *Believability.* Through appearance and behaviors, a pedagogical agent should encourage its perception by learners as a sentient being (cf. Chapter 3.1.3). In particular, Baylor and Ryu (2003) suggested that the key features contributing to a pedagogical agent persona include its tendency to be:
  - *Engaging*. Interactions with the agent should be fun and hold the learner's interest.[28]
  - *Person-like*. The agent should communicate emotions and a personality.
  - *Credible*. The learner should develop confidence in the agent as a trustworthy (cf. Chapter 3.2.2), competent (cf. Chapter 12), and behaviorally consistent (cf. Chapter 7.5) partner.[29]

---

[28] The *engagingness* of educational software or pedagogical agents can be assessed in terms of learners' experiences and attitudes, including their enjoyment of and perceived involvement in the interaction, as well as by looking at related measures of learner behavior, such as time spent in the environment, observed degree of activity, and willingness to use the program or agent again (Gulz & Haake 2006: 323). Gulz and Haake summarized the relationship between attitudinal and behavioral measures as follows: "If a program is found engaging – that is experienced as involving, interesting or having impact – it is likely that users will become more active, stay on longer, and produce more. Engagingness in this sense is not to be equated with *entertainment*" (p. 323, their emphasis). Bickmore et al. (2005) called for making agents "continually engaging to use" in order to encourage their regular and sustained use.

[29] Credibility is not an inherent quality of people, objects, or information but arises as the result of a subjective evaluation of these entities (Fogg 2003: 122). However, according to Fogg, perceived

    - *Instructor-like*. The agent should model behaviors of an instructor or mentor (cf. Chapter 12.3) and stay in character during the interaction.

- *Deictic believability.* Giving advice, explanations, or demonstrations requires pedagogical agents to make natural and unambiguous reference to objects in the learning environment using combinations of (spoken) language, gestures, and locomotion (Lester et al. 1999b; Lester et al. 2000: 124). For this purpose, they have to keep track of the locations of objects in their environment, their own position relative to these objects, and their own previous explanations (Lester et al. 2000: 124).
- *Explanation generation.* Explanations delivered by pedagogical agents have to be situated in the current problem-solving context (Lester et al. 1997a; Lester et al. 2000: 126f.). That is, they should communicate advice or domain knowledge using combinations of media and agent behaviors (speech, gesture, and locomotion) that are appropriate for the current task and situation (Lester et al. 1997a; Lester et al. 2000: 127).
- *Knowledge acquisition, representation, and use.* Pedagogical agents require substantial knowledge about the domain of instruction (Johnson 1998). This knowledge has to be acquired from different sources, manually by human knowledge engineers or automatically using machine learning techniques (cf. Chapter 12.1.1). The acquired knowledge then has to be formalized in some kind of knowledge representation (cf. Chapter 12.1.2) which provides sufficient flexibility and depth to support the agent's instructional interactions and facilitates processes of acquisition and authoring (Johnson 1998). The agent applies its knowledge representation to solve problems related to the domain of instruction, the interaction with the learner, and its own persona. This involves setting (instructional) goals, defining problems to be solved (by the agent and/or the learner) that arise from these goals, finding satisfactory solutions to these problems, and finally using the generated solutions to demonstrate the sequence of problem-solving steps, to coach the learner while he or she is working on a problem (cf. Chapter 2.2.3), or to solve a problem together with the learner (Russell & Norvig 2003: 60f., cf. Chapter 12.1.3).
- *Mixed-initiative interaction.* Agent-learner interactions can be initiated and controlled by the agent, the learner, or both. The first two options are too restrictive and too unpredictable, respectively. However, the third, *mixed-initiative* approach allows the initiative to shift frequently in agent-learner interactions, which gives the learner the opportunity to engage in active problem solving while the agent can still intervene where appropriate (Lester et al. 1997a; Lester et al. 1999a, cf. Chapter 11.3.3.2.1). In general, pedagogical agents should not deprive learners of their sense of control over their actions. Learners might not take the feeling of lack or loss of control too well; furthermore, active learning, where learners are active participators in the learning process (cf. Chapter 2.2.2 and Chapter 2.2.3), is unlikely to take place if the learner's role is reduced to the one of a passive recipient.
- *Instructional expertise.* A pedagogical agent in a teaching or tutoring role requires the ability to present content, to solve problems on its own and explain solutions and problem-solving steps, to diagnose learner behaviors and mistakes, to assess the learner's understanding of the topic, to adapt subject matters and teaching methods to the learner, and to communicate effectively with the learner (cf. Chapter 2.3.3).

---

credibility involves two key factors: perceived trustworthiness and perceived expertise (pp. 122ff.). The trustworthiness dimension indicates the extent to which something or someone is perceived as truthful, fair, and unbiased (p. 123). Expertise (cf. Chapter 12) is the knowledge, experience, and skill perceived in a source (p. 124).

- *Social intelligence.* The challenge for pedagogical agents in interactions with learners is to say or do the right thing in the right way to or with the right person in the right situation at the right time. To address this multi-dimensional challenge, pedagogical agents require social intelligence that enables them to (Johnson et al. 2003b):
    - Carefully time and design their interactions with learners;
    - Respect sociocultural dos and don'ts and exploit cultural stereotypes;[30]
    - Adapt their interactions to the changing cognitive and emotional state of the learner;
    - Communicate understanding and support;
    - Work to build and maintain a long-term relationship with the learner.
- *Student monitoring and modeling.* Assisting and instructing learners properly requires a pedagogical agent to keep track of learners' activities, their overall progress, their cognitive and emotional state, and the changing state of the learning environment (Johnson et al. 2000). What information the agent can collect about the learner and the environment depends on the array of internal and external sensors available to the agent. The former detect changes and learners' actions within the learning environment (e.g. dislocated objects, navigation paths, and problem-solving steps) while the latter monitor aspects of learner behavior occurring outside the environment (e.g. facial expressions, gaze patterns, and speech).[31] Since external sensors are harder to implement and need to deal with more ambiguity and variability in the input signal, the number and diversity of external sensors available to pedagogical agents are often limited. Ideally, the results of monitoring are incorporated into an evolving model of the learner, which allows the agent to adapt its assistance to the learner's situation and needs (cf. Chapter 12.2).
- *Trustworthiness.* Learners must be willing (and able) to trust the information and guidance provided by a pedagogical agent, otherwise they will be reluctant to accept the agent's assistance (Cassell & Bickmore 2000: 50). Since trust is not an instant relationship but can take time to develop, pedagogical agents need to invest time and effort in winning the learner's trust and eventually their friendship (Stronks et al. 2002). To facilitate the relationship building process, the behavior of pedagogical agents should be both believable and predictable (cf. Chapter 3.2.2).

---

30 On the other hand, pedagogical agents also give developers the opportunity to work against negative effects of stereotyping by designing agents in a way that disrupts learners' stereotypical expectations of roles and role behaviors and shows them new perspectives, for example with respect to gender role assignment. However, interfering with the role and value systems of other cultures based on the (implicit) belief that one's own culture (or even one's personal values) are superior in certain respects can be dangerous and may result in strong negative responses from the users of the target culture. For further discussion, see Chapter 7.15 and Chapter 9.2. The issue of whether to break with or exploit (visual) stereotypes in interactions between learners and pedagogical agents was also discussed by Haake and Gulz (2008: 9f.).

31 The distinction between internal and external sensors is most apparent in interactive learning platforms for current computer desktops or the current World Wide Web. It is blurred in virtual reality simulations (e.g. Rickel & Johnson 1999; Rickel & Johnson 2000; Rickel & Johnson 2003) where learners are immersed in the environment, wearing special equipment to allow the tracking of their body movements.

4.3 Summary

This chapter defined pedagogical agents as embodied interactive software agents whose mission is to facilitate human learning by working with students in computer-based learning environments. The first section discussed the capabilities of pedagogical agents, which include monitoring the learner and intervening when necessary, adapting to changes in the learner and the learning environment, giving contextualized and customized explanations, performing demonstrations involving bodily actions in the learning environment, engaging learners in multimodal dialogues, supporting both individual and team learning, and increasing learners' motivation.

As discussed in the second section, pedagogical agents with these capabilities have to meet a number of requirements that also apply to software agents in general (autonomy, adaptiveness, proactiveness, responsiveness, and social ability, cf. Chapter 3.1). To be able to communicate effectively with the learner, pedagogical agents additionally require an articulate digital embodiment. Further requirements that were discussed in the section are more specific to the instructional domain of pedagogical agents and include believability (both as an individual and as an instructor or mentor); deictic believability of combined gestures, speech, and locomotion; mechanisms for the generation and delivery of instructional explanations; the capacity to acquire, represent, and use knowledge about the domain of instruction; the ability to participate in mixed-initiative interactions with the learner; expertise in providing instruction to learners; social intelligence to adjust and time actions and messages; means to collect information about and develop a model of the individual learner; and trustworthiness as a guide and source of information.

5 Interactivity

Everything a human does to or with another human can be called an interaction. Human interactions that use media are mediated human interactions. Everything a human does to or with a computer is a human-computer interaction. (Heeter 2000: 75).

in'ter.ac'tive
1. *New technology that will change the way you shop, play, and learn*
2. *A zillion-dollar industry (maybe). (Newsweek 31 May 1993, cover page).*

Interactivity is an overused, underdefined concept. (Heeter 2000: 75).

One important aspect of embodied interactive software agents is their capacity for 'interaction,' or their 'interactivity.' Both terms are ubiquitous, but what exactly do they imply? Many writers attach them as convenient labels to artifacts, exchanges, and experiences, however rarely with more than an intuitive idea of the underlying concepts. Since the terms 'interaction' and 'interactivity' are used throughout this thesis, it seems advisable to examine them in some detail before moving on, in order to better understand their implications for the design of embodied interactive software agents. The first two sections of this chapter deal with interaction and its relation to communication, respectively. Section 5.3 then provides an in-depth discussion of different views of interactivity that have been proposed in the literature, identifying essential elements of interactive experiences that are useful for agent designers to keep in mind. In Section 5.4, interaction and interactivity are discussed in the context of teaching and learning, with a focus on education taking place at a distance with support from technology (i.e. e-learning, cf. Chapter 2). Finally, Section 5.5 provides a discussion and a summary of the chapter.

5.1 Interaction

Interaction is a fundamental concept in various academic disciplines, among them chemistry, computer science, linguistics, media studies, medicine, physics, psychology, sociology, and statistics. While each of these disciplines has its own specialized view of the concept, in the most general terms, interaction means 'exchange,' 'interplay,' or "mutual influence" (Jensen 1998: 188). More formally, it can be defined as "reciprocal action or influence" (OED 1998: 950) between two or more entities (artifacts, events, organisms, substances, theories, etc.), where the action of each entity may have an effect on and elicit a reaction from the others. Action and reaction are also highlighted as key elements of human interactions with the world in the following definition by Heeter:

An interaction is an episode or series of episodes of physical actions and reactions of an embodied human with the world, including the environment and objects and beings in the world. (Heeter 2000: 80).

Interaction may be immediate (unaided) or rely on technology, by using some kind of intervening medium or interface (cf. Chapter 1.2). Actions and reactions of the participants may or may not be separated in space or time. If the interaction consists of a sequence of exchanges, later exchanges typically relate to and depend on earlier exchanges and the relationships between them (Rafaeli 1988). The capacity for reciprocal action or influence is called *interactivity*. Interactivity is a characteristic (for many even a feature) of technologies, communication settings, or users (depending on the researcher's viewpoint), whose meaning seems intuitively clear. However, interactivity as a concept is surprisingly complex below the surface, as the discussion in Section 5.3 will reveal.

5.2 Interaction and Communication

Interaction is closely related to another type of exchanges, *communication*, and the two terms are often used interchangeably. In fact, in line with a long tradition of previous research, Rafaeli and Sudweeks (1997) contended that "communication is mostly about and for the purpose of interaction." Kress and van Leeuwen also emphasized the close relationship (even dependence) between communication and interaction, by stating that there is no communication without interaction (Kress & van Leeuwen 2001: 114). Communication has been conceived as a form of interaction, involving the purposeful reciprocal exchange of information (messages) and the negotiation of a shared understanding between participants according to a common system of rules and symbols (e.g. natural language):

> *Communication is a form of interaction between agents belonging to a socially sophisticated species; in particular, it is an overt attempt to reach a situation which be [sic!] relatively satisfactory to all the participants. (Tirassa 1997).*

However, not all kinds of interaction involve communication. For example, human operation (or *control*) of computers and other devices involves forms of interaction which are at odds with the reciprocity and negotiation that characterize communication. "Within informatics then, (…) it is possible to have (human-machine) interaction without having communication, but not (computer-mediated) communication without also having (human-computer) interaction" (Jensen 1998: 190). On the other hand, there are forms of communication that do not involve interaction. For instance, broadcasting media like traditional radio and television provide one-way communication from a sender to a group of receivers, with little opportunity for receivers to respond to the broadcasts.

Artifacts can play two roles in interactions involving humans: people may use them to communicate with other people or the artifacts themselves may serve as partners in the interaction. Examples of artifacts in the former role as *medium* include computers, the telephone, and even paper and pencil. In general, *mediation* can be defined as "the interposition of an electronic or mechanical medium by which messages are transmitted between actors" (Burgoon et al. 2002: 661). While various artifacts qualify for use as a medium, the term *mediated interaction* is often reserved for settings that use computers to transmit messages. The use of computers to mediate communications between people is studied by the field of *computer-mediated communication (CMC)* (cf. Chapter 2.3.8). Compared to non-mediated face-to-face interaction between people, which is rich in terms of verbal and non-verbal cues (cf. Chapter 10.1), mediated interaction is characterized by a communicative inventory that is reduced to those cues that can be handled by the chosen medium, in computer-mediated communication for the most part still written text (Burgoon et al. 2002: 661).

Apart from using artifacts to communicate with others, users can also interact with the artifacts themselves, i.e. the artifacts serve as interaction partners to users. The interaction with (and through) artifacts takes place via *user interfaces*. The most important types of user interfaces to computer-based systems were discussed in Chapter 1.2. Interaction with computers is the focus of human-computer interaction (HCI) research (Jacko & Sears 2003).

Interaction through, but also increasingly with, artifacts takes the form of conversations. *Conversation* is a (mediated or non-mediated) linguistic interaction between two or more parties, which can be characterized as (Brown & Levinson 1987; Levelt 1995: 29ff., cf. Chapter 11.3.1):

- *Highly contextualized*. Speakers' contributions must be tuned to those of the other interlocutors, and they need to be anchored in the shared here and now of the interaction.
- *Intentional, goal-directed*. Conversations serve to realize certain communicative intentions of the participants, such as giving information, sharing feelings, making agreements, etc.
- *Cooperative*. Participants mutually assume that their contributions are purposeful, well-conducted, or more generally, cooperative.
- *Polite*. Speakers use language that is geared toward mitigating potential threats to the 'face' (i.e. the public self-image) of their audience.

5.3 Interactivity

Like interaction, the related concept of *interactivity* is studied by researchers from various backgrounds. As a result, the term has different meanings for different people in different contexts, such as communication, e-learning, entertainment, human-computer interaction, information studies, marketing, etc., and a variety of definitions can be found in the literature (e.g. Rafaeli 1988; Blattberg & Deighton 1991; Steuer 1992; Deighton 1996; Rafaeli & Sudweeks 1997; Ha & James 1998; Jensen 1998; Schultz 1999; Aoki 2000; Downes & McMillan 2000; Heeter 2000; Yacci 2000; Liu & Shrum 2002; Kiousis 2002; McMillan 2002; Bucy 2004b; Stromer-Galley 2004). In the early 1990s, interactivity became a buzzword and use of the term exploded in the popular, trade, scholarly, and technical press and media (McMillan 2002). However, the general problem with buzzwords is that "within a certain topic, [they] appear to refer to something very important and (…) – for a given time – are heard constantly, but are often difficult to understand since in reality nobody seems to know what they mean" (Jensen 1998: 185). In this respect, interactivity is similar to another buzzword that also had its heyday during the 1990s and is commonly mentioned in the same breath: *multimedia* (cf. Chapter 2.3.7). We cannot define neither of the two conclusively (yet), but we (think we) know them when we see (or, more broadly, experience) them.

While its definition may present a challenge, most people researching or selling interactivity explicitly or implicitly assume that it is a characteristic of artifacts, exchanges, situations, and relationships that is invariably desirable and commonly has positive effects[32] on a wide range of user-related aspects including acceptance, satisfaction, learning, and mastery; thoughtfulness, cooperation, and responsibility; and performance, motivation, and sociability (Rafaeli 1988; Bucy 2004b: 374).

To shed some light on the 'overused,' 'underdefined,' 'undertheorized,' and 'elusive' concept of interactivity (Heeter 2000: 75; Bucy 2004b: 373), which seems to deconstruct itself under close examination (Rose 1999: 48), various attempts to define it from different perspectives are discussed in the following. The goal of this discussion is to identify elements in these definitions which are helpful for understanding interactivity in the context of interactions between humans and embodied interactive software agents.

[32] However, a growing number of researchers are more reluctant to ascribe unconditional positive effects to interactivity. Some authors have already written about limitations and undesirable consequences of interactive processes (e.g. Liu & Shrum 2002; Sundar et al. 2003; Bucy 2004b).

5.3.1 The Role of Feedback

Interactivity research has its roots in *cybernetics* (Kiousis 2002: 359), which was originally defined by the mathematician Norbert Wiener (1894–1964) as the field of study concerned with "control and communication in the animal and the machine" (Wiener 1948). Modern views of cybernetics extend the original definition to encompass "the study of communication and control within and between humans, machines, organizations, and society" ([INT 37]). One distinguishing feature of cybernetics, compared to Shannon and Weaver's communication model,[33] which became available around the same time (Shannon & Weaver 1949), is its incorporation of *feedback* (Kiousis 2002: 359). In cybernetics, feedback refers to the process in which part of the output of a system is fed back into the system as part of its input in order to regulate its further output. The resulting loop is called the *feedback loop* ([INT 38]). Applied to communication, the ability of the recipient of a message to send a response which influences the sender's next message, has become a principal component of many conceptualizations of interactivity (Kiousis 2002: 359), and the capacity of a medium or communication experience for inducing feedback is commonly viewed as a major criterion for calling that medium or experience 'interactive' (p. 367). This framework locates interactivity in the channel through which communication occurs as a dynamic, interdependent process between senders and receivers (p. 359). The definition by Straubhaar and LaRose illustrates this view:

> *We will use the term* ***interactive*** *to refer to situations where real-time feedback is collected from the receivers of a communications channel and is used by the source to continually modify the message as it is being delivered to the receiver. (Straubhaar & LaRose 1996: 12, their emphasis).*

5.3.2 Interactivity as a Communication Concept

The 1980s and the 1990s saw the rise of the so-called *new media*[34] (Wardrip-Fruin & Montfort 2003), which are nowadays exemplified by web sites (including weblogs and wikis, cf. Chapter 2.3.9), e-mail, chat rooms, virtual reality, streaming technologies for audio and video, CD-ROMs and DVDs, mobile computing, and Internet telephony. Interactivity is generally regarded as a defining component of these and other new media technologies (Sims 1999; van Dijk 1999; McMillan 2002). As a result, further explication of the concept in the context of the new "interactive media" became necessary. Williams et al. offered the following definition:

---

33 The Shannon-Weaver model (Shannon 1948a; Shannon 1948b; Shannon & Weaver 1949) is a general model of communication developed by Claude Shannon and popularized by Warren Weaver, which proposes that there are six common elements to all communication: an information source, an encoder, a message, a transmission channel, a decoder, and a receiver ([INT 39]). Communication is viewed as a one-way process, in which the message is encoded for transmission at the source, transmitted over the channel as a signal, which may be distorted by noise, and decoded by the receiver ([INT 39]).

34 'New media' is a generic term used to loosely describe those forms of media which are made possible through the use of digital technology. They are often contrasted with 'old' media forms originating in the pre-digital era, including analog radio and television as well as printed materials, such as books and magazines.

[T]he degree to which participants in a communication process have control over, and can exchange roles in, their mutual discourse is called interactivity. (Williams et al. 1988: 10, cf. Rogers 1995: 314).

Important aspects of this definition include the notion of control and the need for communication roles to be interchangeable (see below and Rafaeli (1988: 111)). It further indicates that interactivity is not an absolute, present-or-absent characteristic. In fact, different interactive processes may involve different *degrees,* or *levels*, of interactivity. Therefore, interactivity is more appropriately conceived as a continuum or scale rather than an on-off dichotomy:

We must conclude that the point is not: interactivity yes or no. The point is: more or less. All the named characteristics of interactivity are gradients. (Jaspers 1991: 22).

The term “mutual discourse” in the definition by Williams et al. refers to the degree to which a given communication act is based on the sequence of communication acts preceding it. Furthermore, the emphasis of the definition has shifted away from the communication channel towards the relationships between the messages exchanged (Kiousis 2002: 359). This shift is also evident in Rafaeli’s now classic definition:

Formally stated, interactivity is an expression of the extent that, in a given series of communication exchanges, any third (or later) transmission (or message) is related to the degree to which previous exchanges referred to even earlier transmissions. (Rafaeli 1988: 111).

For Rafaeli, interactivity is “quintessentially a communication concept” (Rafaeli 1988: 113), a “process-related, variable characteristic of communication settings” and “not a characteristic of the medium” (Rafaeli & Sudweeks 1997), a view that is not shared by everyone (e.g. Stromer-Galley 2000). According to Rafaeli, “interactivity should apply to a wide range of communication settings: from the unmediated, face-to-face, and intimate to the relatively anonymous and mass mediated” (Rafaeli 1988: 111). His own work has focused on interactivity in computer-mediated communication between humans (Rafaeli & Sudweeks 1997).

Rafaeli’s recursive, message-based definition states that interactivity is “the extent to which messages in a sequence relate to each other, and especially the extent to which later messages recount the relatedness of earlier messages” (Rafaeli & Sudweeks 1997). Full interactivity only begins with the third message in an exchange, when the initiator has the opportunity to incorporate references to the previous exchange in his or her new message. The term *third-order dependency* is commonly used in the literature to refer to the interconnected relationships (or functional coherence) among the messages exchanged (Kiousis 2002: 359; Warnick et al. 2005). Third-order dependency occurs if a person sends a message to another person (action), the receiver replies with a message of his or her own (reaction), and the original sender then reacts to the receiver’s reaction, relating both to the initial message and its response (Koolstra & Bos 2006). Only communication that exhibits such a reciprocal exchange of messages (i.e. third-order dependency) can be called interactive rather than just reactive (Kiousis 2002: 359).

5.3.3 Interactivity as a Two-Way Process

In Rafaeli’s explication, “interactivity varies along a continuum” (Rafaeli 1988; Rafaeli & Sudweeks 1997). Depending on how much later messages relate to earlier messages and their relationships, different levels of interactivity can be distinguished (cf. Figure 18). At the low end of the continuum, there is *one-way (non-interactive)* communication (e.g. broadcast radio

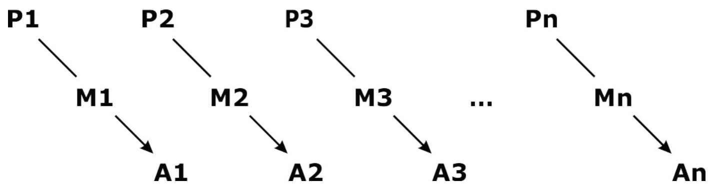

Figure 18a. Non-interactive communication. Adapted from Rafaeli and Sudweeks (1997, Figure 1). (P = Person; M = Message; A = Audience).

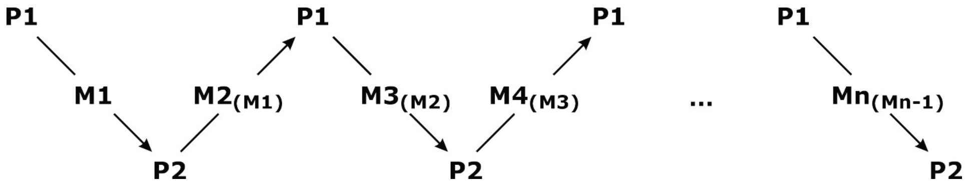

Figure 18b. Reactive communication. Adapted from Rafaeli and Sudweeks (1997, Figure 1).

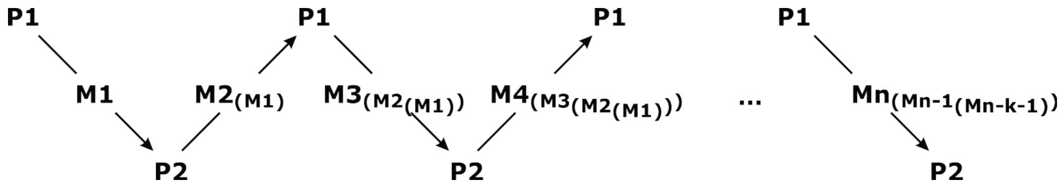

Figure 18c. Interactive communication. Adapted from Rafaeli and Sudweeks (1997, Figure 1).

and television), where messages are not related to their predecessors. *Two-way (reactive)* communication involves one side responding to the other, but only to a single previous message. Finally, a communication is fully *interactive* if a given message refers to at least two earlier messages and the relationship between them (Rafaeli & Sudweeks 1997; [INT 40]). Rafaeli (1988: 119f.) discussed several implications of this model: (a) communication is not always interactive; (b) interactivity is not a characteristic of the medium or channel; (c) much use of new communication technologies is non-interactive; (d) interactivity is distinct from, actually a subset of, feedback. "Interactivity is feedback that relates both to previous messages and to the way previous messages related to those preceding them" (p. 120).

Rafaeli's model involves at least two parties that exchange and build on related messages (Tremayne & Dunwoody 2001: 115). Most researchers share this view that interactive communication is a *two-way* process in which two or more participants function as intentional senders and recipients in a reciprocal exchange of messages (p. 115):

> *Interactivity means two-way communication between source and receiver, or, more broadly multidirectional communication between any number of sources and receivers. (Pavlik 1998: 137).*

Likewise, Markus defined interactivity as a feature of technologies for multidirectional communication (Markus 1990). Many other authors have echoed Pavlik's view (see the quotations in McMillan (2002)). In addition, the *interchangeability* of the roles of sender and receiver is often seen as an important characteristic of interactive exchanges:

> *[F]ully interactive media imply that the sender and receiver roles are interchangeable. (Rice 1984b: 35).*

[T]he degree to which participants in a communication process have control over, and can exchange roles in, their mutual discourse is called interactivity. (Williams et al. 1988: 10, cf. Rogers 1995: 314).

5.3.4 Responsiveness

According to Rafaeli, interactivity resides in the relatedness of messages exchanged among participants in a communication setting (Rafaeli 1988; Rafaeli & Sudweeks 1997). A similar focus on message exchange can be found in the definitions of other authors:

The essence of interactivity is exchange. (Haeckel 1998: 63).

[Interactivity is] the extent to which the communicator and the audience respond to, or are willing to facilitate, each other's communication needs. (Ha & James 1998: 461).

An interactive communication involves responsiveness of the displayed message to the message receiver. (Miles 1992: 150).

*The term **interactive**, as we interpret it, points to two features of communication: the ability to address an individual and the ability to gather and remember the response of that individual. Those two features make possible a third: the ability to address the individual once more in a way that takes into account his or her unique response... (Deighton 1996, his emphasis).*

While it is often perceived as a characteristic of dialogue, interactivity is limited neither to two people nor to face-to-face communication. It can be seen as a variable of responsiveness in interpersonal and societal communication. (Schultz 1999).

A key element of these definitions is the notion of *responsiveness*, which also plays an important role in characterizing interactivity in the context of human interactions with computers: An interactive computer system accepts and responds to user actions (Liu & Shrum 2002), in particular data and commands, while the program is running. Interactive software, such as word processors, spreadsheets, and web browsers, is usually distinguished from non-interactive programs like compilers[35] or batch processing applications,[36] which operate on pre-loaded data without interacting with users ([INT 40]). After the first interactive computer systems had been introduced, "[i]nteractive (…) came to signify a modern, radically improved technology, usually in relation to an older one" (Aarseth 1997: 48), such as batch processing.

In communication, according to McMillan (2002), *responsive dialogue* is characterized by the message sender's awareness of the complete sequence of previous messages as well as his or her control over the exchange. The responsive dialogue model is implemented, for example, by web sites for customer service or e-commerce. In contrast, communication technologies such as chat rooms or instant messaging tools enable *mutual dialogue*, which is responsive but allows all participants to control the exchange. This, in turn, blurs the distinction between sender and receiver roles (McMillan 2002).

---

[35] A *compiler* is a computer program that translates from one computer language into another. The most common application of compilers is to translate (compile) programs written in a higher-level programming language into machine code for direct execution by the computer processor ([INT 41]).

[36] *Batch processing* involves the execution of a series of jobs (consisting of programs and data) on a computer without human intervention ([INT 42]).

5.3.5 Control

Control is another central component of a number of definitions of interactivity. In particular, the field of human-computer interaction (HCI) studies how humans control computers and other new media (McMillan & Hwang 2002: 30). In this research tradition, interaction is "the process that takes places when a human user operates a machine" (Jensen 1998: 190). The concept of control executed by the user over the machine is at the core of the interaction process. At a workshop on the "Methodology of Interaction" in 1979, leading HCI researchers formulated a definition that emphasizes the central role of control:[37]

> *Interaction is a style of control and interactive systems exhibit that style. (Guedj et al. 1980: 69).*

The idea of user control is also emphasized in later technology-based definitions of interactivity, both in the context of human-computer interaction and computer-mediated communication between people, in particular control over form and content:

> *[T]he degree to which participants in a communication process have control over, and can exchange roles in, their mutual discourse is called interactivity. (Williams et al. 1988: 10, cf. Rogers 1995: 314).*

> *[Interactivity is the] quality of electronically mediated communication characterized by increased control over the communication process by both sender and the receiver; either can be a microprocessor. (Neuman 1991: 104).*

> *[Interactivity is] the extent to which users can participate in modifying the form and content of a mediated environment in real time (Steuer 1992: 84).*

> *In interactive systems, a customer controls the content of the interaction requesting or giving information. (Bezjian-Avery et al. 1998: 23).*

> *[I]nteractivity may be defined as: a measure of a media's [sic!] potential ability to let the user exert an influence on the content and/or form of the mediated communication. (Jensen 1998: 201).*

> *We define interactivity as a characteristic of a medium in which the user can influence the form and/or content of the mediated presentation or experience. (Lombard & Snyder-Duch 2001).*

As these definitions indicate, control is not only important for users' interaction with technology but also for interactive communication between users that is mediated by technology. Liu (2003) argued that in both cases, users should have control over both the information sent and the information received to make sure that the reciprocal exchange is satisfactory for all parties involved in the communication. Reciprocal communication provides the channel for exercising control (Liu 2003). Recognizing the importance of both control and reciprocal communication for interactivity, Liu and Shrum defined it as

> *the degree to which two or more communication parties can act on each other, on the communication medium, and on the messages and the degree to which such influences are synchronized. (Liu & Shrum 2002: 54).*

---

[37] However, characterizing the relationship between humans and machines as one of control clashes with another pervasive view in the HCI community that sees human-machine interaction processes as widely analogous to communication processes between people (Jensen 1998: 190).

On the basis of their definition, Liu and Shrum proposed three dimensions of interactivity: *active control* (the user's capacity for voluntary and instrumental action that influences his or her experience), *two-way communication* (reciprocal exchange between the communication parties), and *synchronicity* (the temporal distance between user actions and responses). Synchronicity requires that responses are provided in a timely fashion, i.e. it refers to the speed of the exchange, or the rate of the information flow (Kiousis 2002: 363).

5.3.6 The Real-Time Requirement

In addition to two-way communication and user control, *time* is another important dimension of interactivity, which involves the *speed* at which participants can deliver, process, and respond to messages. In particular, both developers and users view speed of response as a critical feature of interactive media (McMillan & Hwang 2002: 33), as illustrated by the definition below, which focuses on the time requirements of the interaction:

> *Interactive speed is a construct that contributes to flow and is based on measures such as waiting time, loading time, and degree to which interacting with the Web is "slow and tedious." (Novak et al. 2000: 29).*

The emphasis on *response time* implies that the exchange of messages should happen quickly in interactive communication (Liu 2003). That is, contributions by one party should elicit fast follow-up responses from other parties (Bretz & Schmidbauer 1983; Alba et al. 1997; Wu 1999; Liu 2003). The "need for speed" plays an important role in many conceptualizations of interactivity, and a number of authors even go as far as to reserve the label 'interactive' for *real-time*, synchronous exchanges:

> *[Interactivity is] the extent to which users can participate in modifying the form and content of a mediated environment in real time (Steuer 1992: 84).*

> *We will use the term* ***interactive*** *to refer to situations where real-time feedback is collected from the receivers of a communications channel and is used by the source to continually modify the message as it is being delivered to the receiver. (Straubhaar & LaRose 1996: 12, their emphasis).*

> *The ideal interactive medium responds in 'real time' to user input; the response or lag time is not noticeable. (Lombard & Ditton 1997).*

> *Interactivity is the condition of communication in which simultaneous and continuous exchanges occur, and these exchanges carry a social, binding force. (Rafaeli & Sudweeks 1997).*

However, the real-time requirement is at odds with asynchronous online communication using, e-mail, voice mail, and discussion forums, among others, which involves an obvious time lag but is still commonly regarded as interactive (Tremayne & Dunwoody 2001: 115; Kiousis 2002: 362).

Speed is a measure of the rate at which information flows between participants in an exchange (Kiousis 2002: 363). However, speed can be harmful without control, so *time flexibility* (Downes & McMillan 2000), i.e. the degree to which participants can influence the rate of the information flow, is crucial to an interactive experience, in particular if the other party is a machine (Kiousis 2002: 363).

5.3.7 Interactivity and (Non-) Mediation

Speed and time flexibility are obvious characteristics of communication that involves people talking to each other in person, or 'face-to-face.' In face-to-face communication, interactivity characterizes "the relationship between two or more people who, in a given situation, mutually adapt their behavior and actions to each other" (Jensen 1998: 188). The participants in this type of non-mediated communication can make use of a variety of verbal and non-verbal cues (voice, gesture, facial expressions, gaze, locomotion, and posture) to regulate the conversation process, contribute content to the ongoing conversation, and convey emotional and cognitive states (cf. Chapter 10.1). Given these qualities, it comes as no surprise that *interpersonal communication*, and in particular face-to-face communication, is regarded by many as the prototype of interactive communication and hence as a gold standard for interactivity (Bretz & Schmidbauer 1983; Williams et al. 1988; Heeter 1989; Kiousis 2002: 363):

> *[Interactivity is] the capability of new communication systems (usually containing a computer as one component) to 'talk back' to the user, almost like an individual participating in a conversation (Rogers 1986: 34).*
>
> *Face-to-face communication is held up as the model because the sender and receiver use all their senses, the reply is immediate, the communication is generally closed circuit, and the content is primarily informal or 'ad lib.' (Durlak 1987: 744).*
>
> *Interactivity is generally assumed to be a natural attribute of face-to-face conversation, but it has been proposed to occur in mediated communication settings as well. (Rafaeli 1988: 110).*
>
> *[I]nteractivity generally refers to the processes of communication that take on some of the characteristics of interpersonal communication. (DeFleur & Ball-Rokeach 1989: 341).*
>
> *Traditional FtF [(face-to-face)] dialogue is the current prototype for highly interactive encounters because it is contingent, unmediated, same-place, same-time, participative, informationally rich and complex, and identified. (Burgoon et al. 2000: 558).*
>
> *On a simple everyday level, interactivity deals with the ability of systems to simulate interpersonal communication, although the features and components of it change with authors. (Kiousis 2002: 367).*

As the definition by Burgoon et al. quoted above indicates, face-to-face communication possesses a number of features that make it highly interactive in comparison to other communication formats. Burgoon et al. proposed a set of structural properties that allow locating different communication formats along a continuum from high to low interactivity. According to their model, face-to-face communication is more interactive than other formats because it is (Burgoon et al. 2000: 557f.):

- *Contingent*. Participants' contributions relate to previous exchanges, resulting in a flow of communication that is sequenced and coherent but also has an element of unpredictability.
- *Unmediated*. No medium is involved in the exchange of messages between the physically co-present participants.
- *Propinquitous (same-place)*. The interlocutors share the same space.
- *Synchronous (same-time)*. The exchange takes place in real time; there is no delay between or temporal decoupling of messages (i.e. no asychronicity).
- *Participative*. All participants both send and receive verbal and non-verbal messages and feedback to the other participants.

- *Modality- and information-rich*. Participants can make use of rich sensor input, including environmental, visual, audio, and verbal information.
- *Anthropomorphic*. The participants are human (or resemble humans to a high degree, such as robots and embodied interactive software agents).
- *Identified*. The participants know who is involved in the communication; participants are not anonymous.

Media facilitating communication that exhibits more features of face-to-face communication are considered more interactive than others (Jensen 1998: 191). The closer computer-mediated communication resembles interpersonal communication, the more interactive it becomes (Ha & James 1998; Liu & Shrum 2002). A similar emphasis on interpersonal communication can be found in work on interactive media design in psychology, which contends that the effectiveness of an interactive medium depends on how closely it can resemble the interpersonal (Leary 1990; Kiousis 2002: 363).

However, adopting face-to-face communication as the ideal may be an unrealistic approach because interactivity as a concept is typically associated with new media technologies and experienced in technologically-mediated environments rather than the non-mediated real world where face-to-face communication takes place (Jensen 1998: 200; Kiousis 2002: 363 + 373). One way to address this problem is to broaden the notion of interpersonal communication to include both mediated and non-mediated experiences (Kiousis 2002: 363). But a number of authors have argued that the concept of interactivity should be restricted to *mediated* interaction between humans or between humans and technology (e.g. Jensen 1998; Kiousis 2002; Bucy 2004b; Newhagen 2004; Stromer-Galley 2004) (Stromer-Galley 2004: 392). The deliberate exclusion of non-mediated face-to-face communication can be seen in several definitions of interactivity:

> *[Interactivity is the] quality of electronically mediated communication characterized by increased control over the communication process by both sender and the receiver; either can be a microprocessor. (Neuman 1991: 104).*

> *[Interactivity is] the extent to which users can participate in modifying the form and content of a mediated environment in real time. (Steuer 1992: 84).*

> *[I]nteractivity may be defined as: a measure of a media's [sic!] potential ability to let the user exert an influence on the content and/or form of the mediated communication. (Jensen 1998: 201).*

> *We define interactivity as a characteristic of a medium in which the user can influence the form and/or content of the mediated presentation or experience. (Lombard & Snyder-Duch 2001).*

> *Interactivity can be defined as the degree to which a communication technology can create a mediated environment in which participants can communicate (one-to-one, one-to-many, and many-to-many), both synchronously and asynchronously, and participate in reciprocal message exchanges (third-order dependency). With regard to human users, it additionally refers to their ability to* ***perceive*** *the experience as a simulation of interpersonal communication and increase their awareness of telepresence. (Kiousis 2002: 372, his emphasis).*

> *Interactivity, first and foremost, should be reserved to describe reciprocal communication exchanges that involve some form of* ***media****, or information and communication technology. (...) [I]nteractivity should not be considered synonymous with social interaction, person-to-person conversation, or face-to-face communication. (...) Limiting the concept to exchanges that are in some way mediated by technology begins to distinguish the term from* ***any*** *form of communication*

$\mathbf{H} \leftrightarrow \mathbf{H}$ $\mathbf{H} \leftrightarrow \mathbf{C}$ $\mathbf{H} \overset{\mathbf{C}}{\leftrightarrow} \mathbf{H}$ $\mathbf{H} \leftrightarrow \mathbf{D}$ $\mathbf{H} \leftrightarrow \mathbf{A}$

Figure 19a–e. Different types of interactive processes (H = Human; C = Computer; D = Document; A = Agent). Based on Tremayne and Dunwoody (2001: 115, Figure 1).

and discourages its wanton application as a universal descriptor of all forms of dialogue. (Bucy 2004b: 375f., his emphasis).

Whether they include or exclude face-to-face communication, many definitions of interactivity are anthropocentric in nature, i.e. they are concerned with interactions in which at least one of the participants involved is human and tend to focus on human behavior in these interactions. The range of interactive scenarios that have been studied includes interactions between humans; between humans and computers; between humans, mediated by computers (and other technologies); between humans and documents;[38] and, recently, between humans and embodied interactive agents (McMillan 2002, cf. Figure 19). As discussed above, a number of scholars do not regard the first scenario as interactive, although the "exclusion of 'pure' interpersonal communication is debatable" (Kiousis 2002: 373).

5.3.8 Interactivity as Process vs. Product

It might seem attractive to find a single conceptualization of interactivity that covers all the different scenarios in Figure 19. However, according to Stromer-Galley (2004), such an effort may be misguided because it obfuscates the nature and effects of interactivity. She argued that the term 'interactivity' covers two separate aspects: interaction between humans mediated by technology and interaction between humans and computers or networked systems (Stromer-Galley 2004: 391). The former can be described as a process of mediated communication between two or more people, in which new messages make reference to previous exchanges (Bucy 2004b: 375) in the sense of Rafaeli's definition (cf. Section 5.3.2), whereas the latter involves human users interacting with (rather than through) a computer or system equipped with the relevant features (Stromer-Galley 2004: 391), including content presented using that machine through a user interface (cf. Chapter 1.2). Stromer-Galley stressed that interactivity in mediated human-human communication processes and interactivity occurring in user interactions with technology are two distinct phenomena. She introduced the term *interactivity-as-process* as a label for research concerned with human interaction *mediated* by technology in the field of computer-mediated communication. In contrast, *interactivity-as-product* refers to work on human interaction *with* technology that is done in the human-computer interaction tradition (Bucy 2004a: 371; Stromer-Galley 2004: 392). Other researchers have essentially the same dual view of interactivity as Stromer-Galley. For example, McMillan (2002) differentiated between *user-to-user* (process) and *user-to-system* (product) interactivity. Similarly, Lee (2000) identified two broad types of interactivity: interacting with people and interacting with technology. Finally, Carey (1989) provided the following two-part definition of interactive media:

[38] User interaction with documents comprises interaction with both documents and their creators. See McMillan (2002) for a discussion of user-to-documents interactivity.

[Interactive media are] technologies that provide person-to-person communications mediated by a telecommunications channel (e.g. a telephone call) and person-to-machine interactions that simulate interpersonal exchange (e.g. an electronic banking transaction). (Carey 1989: 328).

Given the dichotomy of interactivity-as-process vs. interactivity-as-product, the question arises if it is possible to classify any interactive experience as one of these two types or if there are experiences that have elements of both. In particular, embodied software agents capable of human-like participation in conversations involving the use of the full range of verbal and non-verbal cues available to people in face-to-face situations may be such a 'hybrid' case. On the one hand, software agents are a kind of technology, so the product view of interactivity seems appropriate, which is the one taken, for example, by Bucy:

[Interactivity] can also take the form of ***impersonal interactions*** *with media content or* ***non-human agents*** *– audio/video downloads, e-mail requests to a listserv majordomo, computer game playing, e-commerce transactions, and various other forms of* ***content activity****. (Bucy 2004b: 375, emphasis added).*

But on the other hand, if agent technology can play the part of a human interlocutor in an embodied conversation (cf. Chapter 10.1 and Chapter 11.3), it can be argued that the interaction between humans and embodied interactive software agents moves closer toward interpersonal (even face-to-face) communication while still mediated by technology (the user interface, cf. Chapter 1.2). Hence, it is also possible to consider the human-agent interactive experience as another mediated communication process.

5.3.9 Perceived Interactivity

Scholars may have different views on what interactivity is, but they do not agree on where it resides either. A growing literature is concerned with the issue of locating interactivity in the space defined by users, technologies, and communication settings. The different views of the conceptual whereabouts of interactivity are summarized in Figure 20. The definitions reviewed so far characterize interactivity as a property of the technology (medium) or the communication setting. From the former point of view, interactivity manifests itself in the system's inventory of interface actions (e.g. Steuer 1992), whereas the latter approach observes interactivity in message exchanges, as described in the discussion of Rafaeli's definition above (Bucy 2004b: 376). While technological and message-related accounts focus on objective, observable aspects of interactivity, a third perspective has emerged since the late 1990s, which approaches interactivity from the *subjective* point of view of a participant, suggesting that interactivity is "in the eye of the beholder" (McMillan 2000; McMillan 2002), i.e. it resides, at least in part, in the *perceptions* of individuals (McMillan 2002, cf. Newhagen et al. 1995; Wu 1999; Wu 2000; Lee 2000; McMillan 2000; Downes & McMillan 2000; McMillan & Hwang 2002; Newhagen 2004; Sohn & Lee 2005).[39] By defining interactivity as a perceptual variable (Bucy 2004b: 376), this approach locates the concept *within the user* (Newhagen 2004: 395) rather than the medium or the communication setting. The following quotations from the literature illustrate this view of interactivity as the subjective perception and experience of an individual user:

The experience of interactivity is variant across people and without measuring how people perceive the interactivity, it is difficult to understand the essence of interactivity. (...) [P]revious

[39] See also the discussion of perceptible qualities of agents in Chapter 7.2.

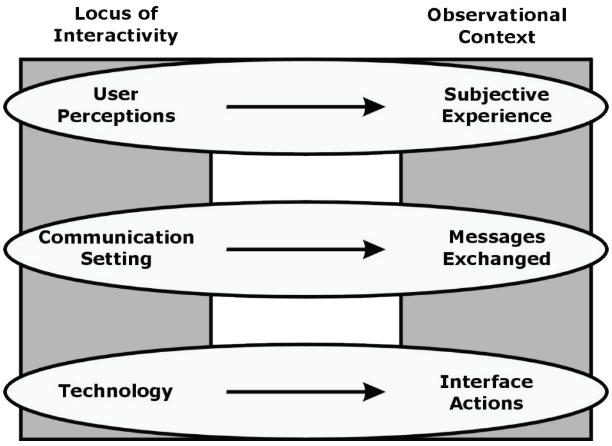

Figure 20. Different views of where interactivity resides and where it can be observed. Adapted from Bucy (2004b: 376, Figure 1).

> *research clearly shows that the most important to be examined in measuring the level of interactivity is not counting mere provision of technological features but investigating how users perceive/experience. User perception of interactivity is the key variable to be examined. (Lee 2000).*
>
> *[Interactivity is] the extent to which a person perceives he or she controls over [sic!] the interaction process [and the extent to which the person perceives that] his or her communicative counterpart (a person, a mass-mediated environment, or a computer-mediated environment) personalizes and responds to his or her communicative behavior. (Wu 2000: 41).*
>
> *Ultimately it is the consumer's choice to interact, thus interactivity is a characteristic of the consumer, and not a characteristic of the medium. The medium simply serves to facilitate the interaction. (Schumann et al. 2001).*
>
> *With regard to human users, (...) [interactivity] refers to their ability to **perceive** the experience as a simulation of interpersonal communication and increase their awareness of telepresence. (Kiousis 2002: 372, his emphasis).*
>
> *Fundamentally, human communication is concerned with **meaning** derived from **content** embedded in physical objects called **symbols** that serve as the base units in narratives, programs, and codes called **messages**. Content creation, provoked by **interaction** between internal mental states, takes place when information is embossed onto symbols. Interactivity, then, is an information-based process that takes place within the individual. (Newhagen 2004: 395, his emphasis).*
>
> *Human perception of interactivity is indispensable in studying the effects of interactive media on individuals: Whether people actually perceive a medium/vehicle as interactive is the only valid criterion for judging its interactivity. (Sohn & Lee 2005).*

Scholars studying interactivity as user perception have criticized technological definitions for their lack of consideration for the *experiential* dimension of interactivity. According to Bucy, regarding interactivity as a technological factor and looking for it only in physically observable actions and reactions (e.g. Heeter 2000; Sundar 2004) fails to account for the subjective qualitative experiences which individuals and different groups of individuals may equate with interactivity (Bucy 2004b: 376). In fact, virtually any mediated setting, including mass communication (which is often said to be non-interactive), may be perceived by users as offering possibilities for interaction or participation, even if it lacks many of the features typically associated with interactivity in the research literature (p. 376). In these media settings, users may not have any real control over the content or perform observable communicative actions, but nonetheless, they may have the *sense* of “participating in the ongoing action of the representation” (Laurel 1991: 20f.) and feel that they are part of an interactive experience (Bucy 2004b: 376). Furthermore, users with different skill levels and experience may perceive different sets of affordances, or sets of possible actions, in the same media technology (p. 376), which implies that on the one hand, it is possible for them to miss features, but on the other hand also that they can use or combine available features to enable new interactive experiences not foreseen by the designers of the technology (cf. Chapter 6.1).

The perceived interactivity of ‘objectively’ non-interactive communication settings indicates the psychological dimension of the phenomenon (Bucy 2004b: 377). Bucy argued that “any networked medium or communication setting *perceived* as interactive becomes so to the user” (p. 377, his emphasis). From their seminal empirical research on the interaction between humans and media (cf. Chapter 6.5), Reeves and Nass concluded the following:

> *[I]ndividuals’ interactions with computers, television, and new media are* ***fundamentally social and natural****, just like interactions in real life. (Reeves & Nass 1998: 5, their emphasis).*
>
> *What* ***seems*** *to be true is often more influential that [sic!] what really is true. Our studies show that* ***perceptions*** *are far more influential than reality defined more objectively. When perceptions are considered, it doesn’t matter whether a computer can really have a personality or not. People perceive that it can, and they’ll respond socially on the basis of perception alone. (Reeves & Nass 1998: 253, their emphasis).*

Similarly, research by Turkle (1984) suggests that humans tend to perceive interactive systems as possessing human characteristics like being ‘alive’ or ‘cheating’ (Kiousis 2002: 364, cf. Chapter 9.1). Given these results, it appears that

> *to some extent, interactivity is associated with the ability of individuals to experience different media as if they were engaging with other human beings. Accordingly, the simulation of interpersonal communication in an interactive environment is not confined to human-to-human communication, but includes human-to-machine communication as well. (Kiousis 2002: 364).*

If this conclusion is correct, it follows that communication between humans and embodied interactive software agents is another kind of experience that users can perceive and understand as interactive. In general, the lesson from research on interactivity as user perception is that interactivity as a concept cannot be understood without considering what happens inside the individual user’s mind, i.e. how he or she perceives and makes sense of an interactive experience.

However, the conceptualization of interactivity as user perception is not without its critics. For example, Sundar contended that “interactivity is an attribute of the technology and not that [sic!] of the user” (Sundar 2004: 385). He argued that situating interactivity in the user rather than the medium invites subjective evaluations of the kind “I know it when I see it” (p. 386). Furthermore, he wrote, evaluations of interactivity would be influenced by the user’s skill level: if the user is skilled enough to use a given interface, he or she would give it fairly high interactivity ratings, whereas a user with a lower skill level would rate it more negatively. The result would be a confusion of perceived interactivity and perceived *usability* (cf. Chapter 6.3) (p. 386).

Another problem with locating perceived interactivity within the user is that it is not physically observable and hence not immediately available for empirical investigation. For the purposes of empirical research, the concept has to be *operationalized*, i.e. associated with a set of measurable units (cf. Chapter 15.4.3). McMillan et al. (McMillan & Hwang 2002; McMillan et al. 2003) presented a set of 18 scaled *measures of perceived interactivity (MPIs)* to operationalize perceived interactivity in the context of measuring consumer perceptions of advertising on the World Wide Web. These measures are organized in terms of three subdimensions (McMillan et al. 2003: 402):

- *Real-time communication*. Capacity for conversation, two-way communication, concurrent communication, and interpersonal communication;
- *No delay*. Speed of the transaction between human and computer;
- *Engaging*. Content variety, appeal to the user, ease of navigation, and immediacy of response.

Recent research (Sohn et al. 2003) has linked the concept of perceived interactivity to the expectations of users. As illustrated in Figure 21, Sohn et al. argued that interaction is a circular process in which users' prior *expected interactivity* with respect to a chosen medium influences their actual interaction with instances of the medium and is, in turn, modified by users' perceptions of the interactivity of that experience. Sohn et al. defined expected interactivity as "the extent of interactivity that a person expects to experience during a prospective interaction with a medium" (Sohn et al. 2003). For example, users' subjective expectations about interactions with the World Wide Web in general *before* they experience an interaction with a web site will affect their actual interactions with specific web sites, which, in turn, will influence their perception of the interactivity of these sites *after* the interaction. Finally, users' post-interaction evaluations of the interactivity of web sites will modify their expectations of interactivity with regard to the World Wide Web as a medium (Sohn et al. 2003). Hence, Sohn et al. proposed two levels of interactivity from the perspective of users: their expected interactivity toward a medium and their perceived interactivity of particular instances of that medium. The study conducted by Sohn et al. (2003) confirmed the role of expected interactivity in understanding the relationship between the user and the medium. In particular, expected interactivity was found to affect perceived interactivity when an instance of a medium (a web site) exhibited high-level interactive features (selected according to the dimensions of interactivity proposed by Ha and James (1998)).

5.3.10 Multidimensional Views on Interactivity

The discussion so far has shown that interactivity is a multi-faceted concept (Downes & McMillan 2000), which is difficult to capture with a definition that focuses on a single aspect, such as message exchange, as in Rafaeli's classic account presented above. In fact, an increasing number of researchers acknowledge that interactivity should be conceptualized along multiple dimensions, with each dimension being represented by a continuum (Downes & McMillan 2000). An early *multidimensional* definition of interactivity by Heeter (1989) placed the concept in the structures and processes of the medium (McMillan 1998). She identified six measurable dimensions of interactivity, which she intended to cover all new (pre-Web) media (Heeter 1989: 225–231; Jensen 1998: 199f.; McMillan 1998; Chou 2003):

1. *Complexity of choice available (selectivity)*. The amount and variety of choices (actions or information) available to the user (this element is often associated with control, cf. Section 5.3.5);
2. *Effort users must exert*. How much physical or mental energy users require to access information or accomplish other tasks;
3. *Responsiveness to the user (conversationality)*. The extent to which a medium can react responsively to a user;

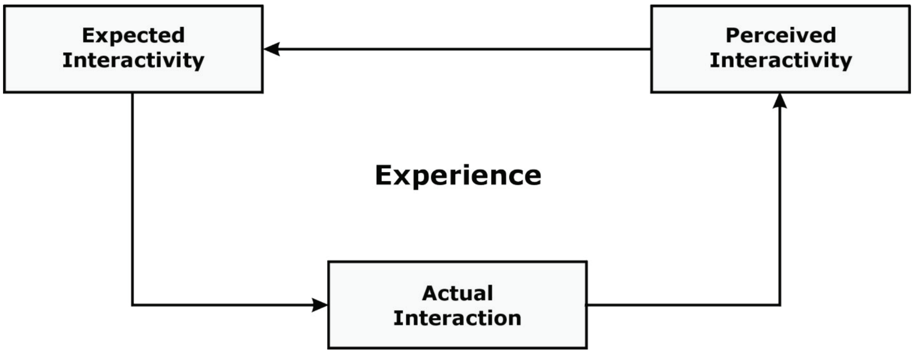

Figure 21. Sohn et al.'s circular process of interactivity perception formation (Sohn et al. 2003, Figure 1). The diagram shows the interplay between prior expectations, interaction, and post-interaction evaluation.

4. *Monitoring information/system use*. How well data about the behavior of users while on the media system can be collected across an entire user population;
5. *Ease of adding information*. The degree to which users can contribute content to a media system that can be accessed by a large, undifferentiated audience;
6. *Facilitation of interpersonal communication*. The extent of support available for (a-) synchronous mediated person-to-person communication.

Heeter's catalogue clearly emphasizes the role of the user (Kenney et al. 2000). Four dimensions (nos. 1, 2, 3, and 5) are concerned with the choices available to the user, what effort he or she has to expend, how responsive the medium is to him or her, and how easily he or she can contribute information, respectively. The sixth dimension treats both parties involved in a communication equally. Finally, the fourth dimension considers the other side of the two-way (or more) flow of information (the sender). Heeter's conceptual definition has been adapted for several empirical investigations of interactivity, however with mixed success (e.g. McMillan 1998; Massey & Levy 1999; Kenney et al. 2000). Jensen saw the problem of manageability when applying Heeter's definition in practical situations, due to the large number of dimensions, their complexity, and the fact that they overlap each other (Jensen 1998: 200).

Ha and James (1998) analyzed the interactivity of business web sites, defining the concept as "the extent to which the communicator and the audience respond to, or are willing to facilitate, each other's communication needs" (Ha & James 1998: 461, see above). To operationalize interactivity, they proposed five dimensions (Ha & James 1998; Kiousis 2002: 367 + 380):

1. *Playfulness*. The entertainment value of an interactive experience, which is provided by features including interactive games and "curiosity-arousal devices" (e.g. humorous animations);
2. *Choice*. The number of options available to users (regarding color, speed, language, etc.);
3. *Connectedness*. "[T]he feeling of being able to link to the outside world and to broaden one's experience easily" (Ha & James 1998: 462); operationalized in terms of the number of hyperlinks on web pages;
4. *Information collection*. The presence of mechanisms for monitoring user behavior (similar to Heeter's fourth dimension);
5. *Reciprocal communication*. The availability of response (feedback) mechanisms for users (e.g. contact forms, surveys, shopping baskets, etc.).

In Steuer's conceptualization, interactivity is determined by the "technological structure of the medium" and defined as "the extent to which users can participate in modifying the form and content of a mediated environment in real time" (Steuer 1992: 84). Steuer examined three factors (or dimensions) that contribute to interactivity (p. 85f.):

1. *Speed of interaction (or response time)*. The rate of assimilation of user inputs enabled by the mediated environment. Real-time interaction (cf. Section 5.3.6) scores high on this dimension, as user actions change the environment immediately.
2. *Range*. The number of possibilities for effecting change on the mediated environment at any given time (determined by temporal ordering, spatial organization, intensity (of sounds, images, and smells), and frequency characteristics (timbre, color)).
3. *Mapping*. The ability of a system to connect human actions (through system controls) to actions in the mediated environment in a natural and predictable (vs. arbitrary and unrelated) manner. Mapping can be increased by adapting controllers to the affordances provided by the human body (e.g. speech recognition, data gloves, eye gaze tracking, etc.).

In an attempt to operationalize interactivity in the context of computer-mediated environments, Downes and McMillan (2000) conducted in-depth interviews with individuals working in the field of interactive communication. From the responses which they received, they identified six dimensions of interactivity (Downes & McMillan 2000):

1. *Direction of communication*. Support for two-way communication provided by the computer-based interactive environment;
2. *Time flexibility*. The extent to which the timing of communication can meet the varying time demands of participants and can be controlled by them;
3. *Sense of place*. The degree to which the environment can make users feel that they have been transported to a virtual place;
4. *Level of control*. Users' perception of the degree to which they can exert control over content, timing, participation, etc. in the environment;
5. *Responsiveness*. The degree to which participants perceive the communication in the environment to be reacting responsively to their own actions;
6. *Perceived purpose of communication*. Individual perceptions of communication goals in the environment (informational vs. persuasive).

Jensen (1998) contributed another multidimensional conceptualization of interactivity, based on his definition of the concept as "a measure of a media's [sic!] potential ability to let the user exert an influence on the content and/or form of the mediated communication" (Jensen 1998: 201). He suggested four dimensions of interactivity, two of which can be conflated into a single dimension (pp. 201f.):

1. *Selective*. The availability of choice, with or without a return channel enabling users to make requests to the media system. This dimension includes two subdimensions:

   - *Transmissional*. The availability of choice in a one-way media system with no return channel and hence without the possibility to make requests;
   - *Consultational*. The availability of choice (by request) in a two-way media system with a return channel;

2. *Conversational*. The extent to which a medium allows the user to create and contribute his or her own information, whether stored or in real time;
3. *Registrational*. The ability of a medium to collect information about a user and to use this information to adapt and/or respond to the user.

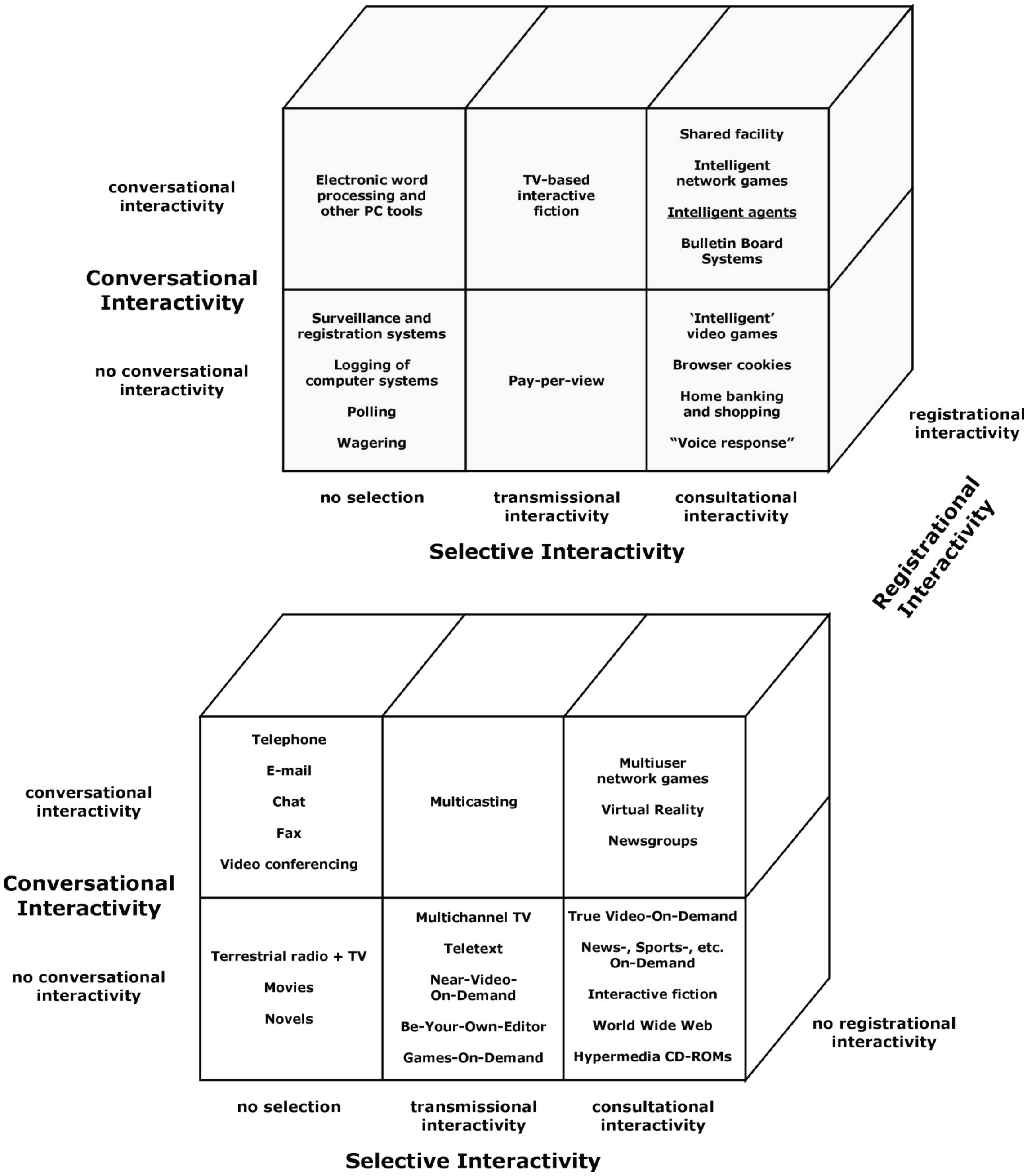

Figure 22. Jensen's three-dimensional representation of the dimensions of interactivity. Adapted from Jensen (1998: 202, Figure 9). For purposes of presentation, the cuboid has been split into two halves orthogonal to the registrational dimension.

Jensen integrated the three dimensions (selective, conversational, and registrational) into a three-dimensional graphical model of interactivity, which is shown in Figure 22. It is worth noting that Jensen's model includes intelligent agents (underlined in Figure 22), which he classified as 'consultational,' 'conversational,' and 'registrational.' That is, in Jensen's model of interactivity, agents are two-way media systems that provide a return channel for user requests, accept different kinds of (multimodal) input from the user, and collect information about each user to adapt their behavior and responses to his or her actions, knowledge, needs, preferences, etc.

5.4 Interactivity in (Distance) Education

No topic raises more contentious debate among educators than the role of interaction as a crucial component of the educational process. This debate is fueled by surface problems in definition and vested interests of professional educators, but is more deeply marked by epistemological assumptions relative to the role of humans and human interaction in education and learning. (Anderson 2003b: 1).

A long history of study has recognized interaction and interactivity as central ingredients of teaching and learning processes and contexts (e.g. Dewey 1916; Daniel & Marquis 1988; Moore 1989; Hillman et al. 1994; Wagner 1994; Sims 1997a; Sims 1997b; Berge 1999; Sims 1999; Sims 2000; Yacci 2000; Bannan-Ritland 2002; Anderson 2003a; Anderson 2003b; Chou 2003; Sims 2003; Wallace 2003; Anderson 2004; Muirhead & Juwah 2004; Phillips 2004; Roblyer & Wiencke 2004; Woods & Baker 2004; Wagner 2006), whether education happens face-to-face (same place and time) or at a distance (independent of place and/or time), with or without support from educational media (cf. Figure 23).

Since the arrival of the new media in (distance) education, the potential to create interactive learning experiences with them has led many to consider learning *from* (as an instructor or tutor), *with* (as a tool), and *through* (as a medium for interpersonal communication with learners, instructors, and other people) information and communication technologies (Yacci 2000) as superior to face-to-face classroom teaching and other traditional forms of instruction. In particular, the various applications of computers in education since the 1960s and increasingly since the 1990s, including, in chronological order, computer-assisted instruction (CAI),[40] multimedia learning, and, recently, e-learning and mobile learning (cf. Chapter 2.1), have all been regarded as superior to traditional instruction because each of them is 'interactive' (Yacci 2000):

Computer-based instruction provides greater potential for truly interactive instruction than any mediated teaching device to date, excluding in many instances, the human tutor. (Jonassen 1988: 97).

*The word **interactive**, when used to describe computer-based learning resources, has tended to imply better experiences, more active learning, enhanced interest and motivation. (Sims 1999, his emphasis).*

[I]nteractivity is frequently used to refer to an inherent quality of the medium and learning environment, with an underlying assumption that the interactive characteristics of communication with other learners or content objects is beneficial to the learning process. (Sims 2003: 87).

For technology-mediated learning, interaction is a key value proposition. (...) Interaction continues to be perceived as the defining attribute for quality and value in online learning experience [sic!]. (Wagner 2006: 44).

While the beneficial effects of interaction and interactivity in education seem to be disputed by few, there is no unanimous definition of these concepts in the educational literature, just as

40 Computer-assisted instruction (CAI) is only one example from the large collection of terms in use to refer to the application of computers and related technologies in education. Apart from CAI, many other abbreviations have been built from the initial letters of words including 'computer' (C), 'managed' (M), 'based' (B), 'assisted' (A), 'aided' (A), 'learning' (L), 'instruction' (I), 'training' (T), 'education' (E), and so on.

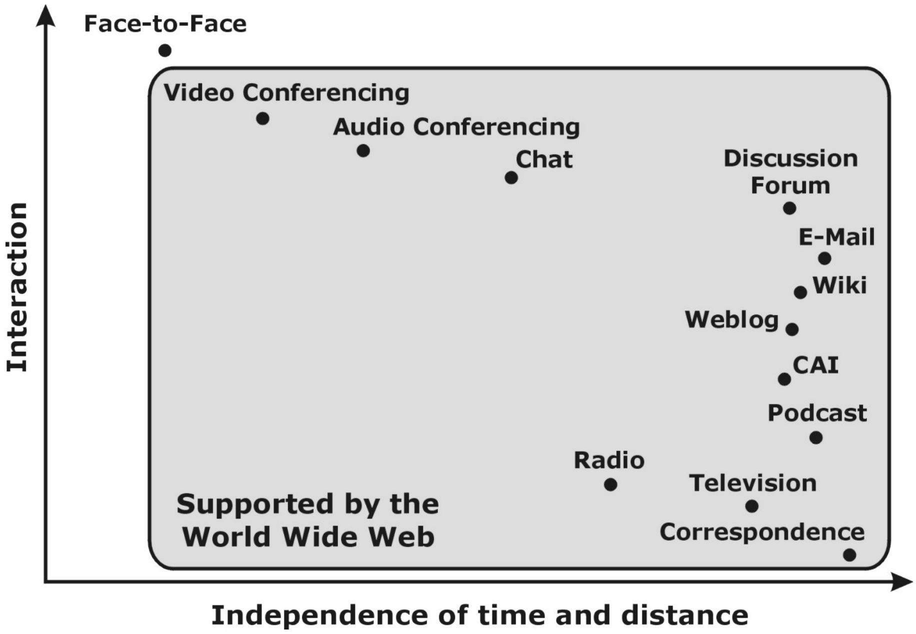

Figure 23. A qualitative diagram of different educational media charted against the degree of interaction and independence (of time and distance) supported by these media. Adapted from Anderson (2004: 44f., Figure 2.1 and Figure 2.2). As the extension of the box indicates, the World Wide Web provides support for all the educational media shown in the diagram.

no one has been able to define them conclusively in general, as discussed in the previous sections of this chapter. The overuse and underdefinedness (Heeter 2000: 75) of both terms in the context of education has been noted by a number of scholars:

> *Interaction is another important term that carries so many meanings as to be almost useless unless specific submeanings can be defined and generally agreed upon. (Moore 1989: 1).*

> *Interactivity is important, but there appears to be no consensus of what interactivity actually represents or involves. (Sims 1997a: 160).*

> *[D]econstruction essentially reveals interactivity to be not a conceptual unity, defined in terms of clear distinctions between antithetical terms, but as a fragmented, inconsistent, and rather messy notion encompassing both privileged and marginalized binaries, and the range of meanings in between. (Rose 1999: 48).*

> *[A]lthough interaction has long been a defining and critical component of the educational process and context, it is surprisingly difficult to find a clear and precise definition of this concept in the education literature. (Anderson 2004: 43).*

To better understand the role of interaction and interactivity in the context of teaching and learning, the following sections discuss functions of interaction and interactivity in education (cf. Section 5.4.1), describe different types of interaction in (distance) education (cf. Section 5.4.2), explore the nature of instructional interactivity (cf. Section 5.4.3), review the concept of transactional distance (cf. Section 5.4.4), examine the contribution of social interaction to educational experiences (cf. Section 5.4.5), and, finally, consider interactivity in the context of learning from computers (cf. Section 5.4.6).

5.4.1 Functions of Interaction and Interactivity

Many researchers and practitioners in the field of education stress that interaction between and among learners, learning content, instructors, and other elements of the educational process and context is important for achieving desired learning outcomes:

> *[Interactivity in learning is] a necessary and fundamental mechanism for knowledge acquisition and the development of both cognitive and physical skills. (Barker 1994: 1).*
>
> *Interaction is intrinsic to successful, effective instructional practice as well as individual discovery. (Sims 1997a: 158).*

Interaction serves a wide range of functions in the educational process. Some examples of the positive effects of interaction include the following (Northrup et al. 2002; Muirhead & Juwah 2004; Mason & Rennie 2006):

- Interaction fosters active and participative learning of individuals and within groups.
- Interactions can be designed to suit different learner needs and learning styles.
- The capacity for interaction enables learners to contribute to and assume ownership and control of their learning process.
- Appropriate interactions can help learners to develop higher-order knowledge and skills, including critical thinking, decision making, problem solving, reflection, etc.
- Interactions give all parties involved the opportunity to provide feedback to each other that informs the learning process and enhances the learning experience.
- Social interactions among participants in a (distance) learning experience are critical for building social presence and community, with positive effects on learning (cf. Section 5.4.5).
- Interaction is important for learner satisfaction and contributes to student persistence in distance learning courses.

Given these benefits, it is not surprising that more, and more sophisticated, interaction (or interactivity) in a learning experience, in particular one that is mediated by technology, is commonly expected to improve learning outcomes and learner satisfaction:

> *The rationale for [exploring interaction in the context of technology-mediated learning] (...) revolves around two commonly held beliefs:*
>
> - *The perceived quality of a learning experience is directly proportional to and positively correlated with the degree to which that experience is seen as interactive.*
> - *If technology-mediated learning designs are to have any significant impact on current and future pedagogical practice, then learning design and development decisions need to maximize the benefit of interaction. (Wagner 2006: 45).*

Interaction and interactivity in education have also become increasingly important because of the high educational value attributed to them in modern theories of learning, which assign an actively constructive role to the learner (cf. Chapter 2.2.3). Looking at the historical progression of learning theories from behaviorism (cf. Chapter 2.2.1) through cognitivism (cf. Chapter 2.2.2) to present-day constructivist theory, there has been a concomitant rise in the expectations of interactivity levels in learning activities. A number of taxonomies in the literature reflect the common view that the higher the level of interactivity, the better the outcome (see below). For example, Rhodes and Azbell (1985) proposed three levels of interactivity relating to the degree of user control (cf. Section 5.3.5) over content and structure. The lowest, *reactive* level involves basic stimulus-response interactions of a behaviorist character, granting learners only limited control over content and structure. The second, *co-active* level gives the learner more control over either content or structure, but not both. Finally, at the highest, *proactive* level, the learner has extended control over both content and structure (Sims 1999). The notion of increased learner control corresponds well with constructivist ideas of learning (Metzger & Schulmeister 2004: 266).

5.4.2 Types of Interaction

In processes and contexts of education, interaction has traditionally been conceived as an exchange between people. For example, Daniel and Marquis understood interaction "in a restrictive manner to cover only those activities where the student is in two-way contact with another person (or persons)" (Daniel & Marquis 1988: 339). Imposing the same restriction, Garrison defined interaction as "sustained, two-way communication among two or more persons for purposes of explaining and challenging perspectives" (Garrison 1993: 16). These definitions imply that in a formal educational environment, interaction is seen as occurring between learners and instructors, or among learners (Berge 1999). However, the participation of other elements in interactions involving learners has also been recognized. Moore (1989) and Juler (1990) described interaction between learners and content as another critical component of education (Anderson 2003b). Moore's now classic trichotomy specified three distinct types of interaction in the context of distance education, defining interactions as transactions between learner and content, learner and instructor, and learner and learner (Moore 1989):

- *Learner-content interaction*. The process of learners examining, considering, and assimilating the instructional content presented during the educational experience. Moore regarded this type of interaction as a "defining characteristic of education" (Moore 1989) because "[e]very learner has to construct knowledge through a process of personally accommodating information into previously existing cognitive structures [(cf. Chapter 2.2.2)]. It is interacting with content that results in these changes in the learner's understanding" (Moore & Kearsley 1996: 128).
- *Learner-instructor interaction*. Communication between the learner and the instructor in which the learner can draw on the assistance, advice, encouragement, information, organization, and other kinds of support provided by the instructor (cf. Chapter 12.3), and the instructor can use feedback (cf. Section 5.3.1) from the learner to adapt to his or her needs. Mutual feedback between individual learners and the instructor is essential for this type of interaction; however, this is not feasible in all educational contexts.
- *Learner-learner interaction*. Communication of learners with one or more peers in which the instructor may be present and/or influence the interaction or not. According to Moore, the value of learner-learner interaction "depends largely on the circumstances of the learners and their age, experience, and level of learner autonomy. For younger learners, the teaching task of stimulation and motivation will be assisted by peer-group interaction, though this is not particularly important for most adult and advanced learners, who tend to be self-motivated" (Moore 1989: 5).

Moore's three-part construct of educational interaction has seen several extensions and adaptations in subsequent, more complex models of interaction in educational settings (Woods & Baker 2004). Hillman et al. (1994) supplied a fourth interaction type, *learner-interface interaction*, acknowledging the increasing importance of user interfaces (cf. Chapter 1.2) in technology-based distance learning environments for facilitating interaction between content, learners, and instructors. Being able to interact successfully with the interface was regarded as a prerequisite for learners to be able to interact effectively with content, learners, and instructors (Godwin 2005). "When dealing with any tool, it is necessary for the user to interact with the device in a specific way before it will do his or her bidding" (Hillman et al. 1994: 34). In other words, the interface is not a neutral component but influences interactions of the other types (Godwin 2005). It can even become "an independent force with which the learner must contend" (p. 35), in particular for less technically-adept learners, and learners may have to be literate with several interfaces (p. 39).[41]

Another type of interaction was found by Burnham and Walden (1997). On the basis of their observations of the interactions in a distance learning environment, they added *learner-environment interaction* to the set of interaction types (Woods & Baker 2004), which they defined as "a reciprocal action or mutual influence between a learner and the learner's surroundings that either assists or hinders learning" (Burnham & Walden 1997). Learners may interact with their environment by conducting experiments, participating in field trips, attending in-class sessions, and doing other activities outside the computer-based learning environment (Godwin 2005).

Three further types of interaction in distance education were added to Moore's (1989) framework by Anderson and Garrison (1998): *instructor-instructor*, *instructor-content*, and *content-content interaction* (cf. Figure 24).[42] The interaction of instructors with peers in communities of colleagues sharing their interests provides them with opportunities for professional development and support (Anderson 2004: 48, cf. Chapter 2.2.3). This type of interaction takes place at conferences, in seminars, through informal communication, etc. (Anderson & Garrison 1998: 105). Instructor-content interaction occurs as instructors create content and learning activities (Anderson 2004: 48). The World Wide Web and other recent technological achievements have enabled instructors to develop, collate, monitor, and update content and activities in more efficient and creative ways than before (Woods & Baker 2004). Finally, content-content interaction was introduced by Anderson and Garrison to acknowledge that more and more content is integrated with increasingly sophisticated tools (Woods & Baker 2004), including databases, intelligent agents, search engines, weblogs, wikis, content syndication aggregators, etc., which enable content to update and improve itself through interaction with other automated resources (Anderson 2004: 48).

In later work, Anderson (2003b) addressed the issue of finding an appropriate mix of the different interaction types in distance education in order to achieve deep and meaningful learning at acceptable costs and expense of time. In particular, he considered the equivalence of interaction modes and its implications for the design of educational experiences. Anderson (2003b: 3f.) noted that:

- The transformation of instructor-learner and learner-learner interaction into enhanced forms of learner-content interaction is both feasible and desired.

[41] Pedagogical software agents implemented as embodied interactive characters can help to 'humanize' the interface to a computer-based learning environment by allowing learners to draw on their pre-existing knowledge about human-human communication in their interactions with an agent. In addition, various forms of human embodied interaction with the environment itself can be considered as alternatives to traditional human-computer interfaces (Dourish 2004).

[42] Anderson and Garrison (1998) used the terms 'teacher' and 'student' rather than 'instructor' and 'learner' in their framework. In the present discussion, the latter terms have been adopted instead to be consistent with the terminology chosen for the other types of interaction.

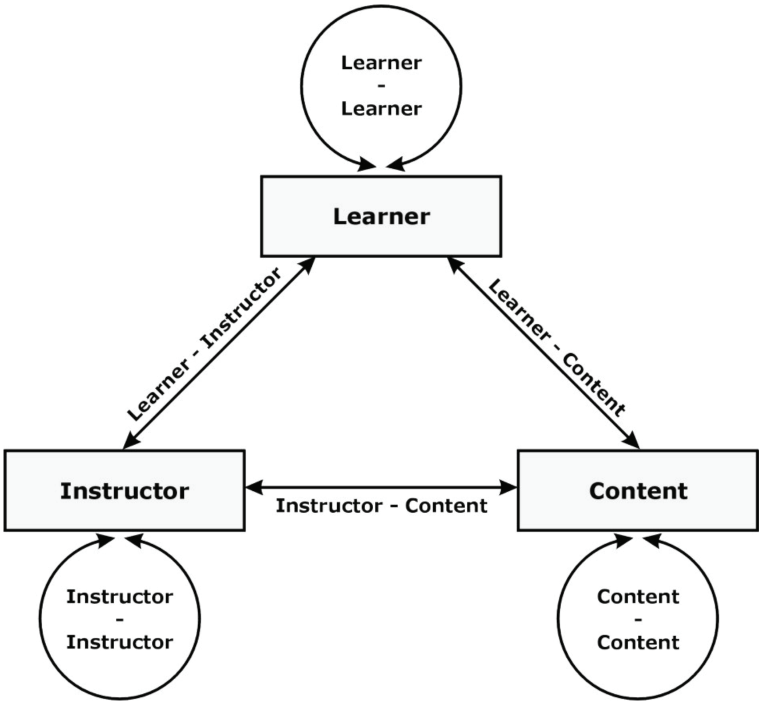

Figure 24. Six types of educational interaction in distance education. Adapted from Anderson and Garrison (1998).

- Media are equivalent in their ability to support educational experiences.
- The impact of media on learning is subject to a complex interaction between various factors, including content, learner preferences and needs, facilities and policies of institutions, and theoretical and practical frameworks of teaching and learning.
- While educational theorists stress the importance of interaction, many learners opt for programs requiring minimal interaction with instructors and other learners.
- "[T]here is a wide range of need and preference for different combinations of paced and unpaced, synchronous and asynchronous activity, and also a strong desire for variety and exposure to different modes and modularities of educational provision and activity" (Anderson 2003b: 4).

Based on these observations, Anderson proposed an *equivalency theorem* to provide a theoretical basis for judging the appropriate relative proportions of learner-instructor, learner-learner, and learner-content interaction in educational settings:

> *Deep and meaningful formal learning is supported as long as one of the three forms of interaction (student-teacher; student-student; student-content) is at a high level. The other two may be offered at minimal levels, or even eliminated, without degrading the educational experience.*[43]

[43] The team behind the Virtual Linguistics Campus (VLC) can actually confirm Anderson's theorem from its own practical experience of both developing the VLC and having used it to deliver web-based courses since 2001. During this time, it has been observed again and again that the high level of learner-content interaction provided by the classes in the VLC reduces the need for learner-instructor and learner-learner interaction, since the content is largely self-explanatory (Handke 2006b: 141).

> *High levels of more than one of these three modes will likely provide a more satisfying educational experience, though these experiences may not be as cost or time effective as less interactive learning sequences. (Anderson 2003b: 4).*

According to Anderson, the central implication of this theorem is the equivalence of the three types of interaction: one type can be substituted for one of the other two without diminishing educational effectiveness, provided that the substitute is at the same (high) level (Anderson 2003b: 4). In the following, he listed further corollaries and implications of his theorem (pp. 4ff.). From the perspective of this thesis, it is interesting to note that one of them is that "[s]ome student-teacher interactions can be *automated*, and thus substituted in whole or part, through the development and use of content resources, and especially those utilizing *autonomous teacher agents*" (p. 4, emphasis added), which assist the instructor. Teacher agents perform functions that support instructors in the administration of the educational process. These functions include (p. 5):

- Communicating with student and content agents;
- Monitoring and modeling learner progress (cf. Chapter 12.2);
- Collecting and organizing educational resources;
- Helping with marking, record keeping, and document management;
- Scheduling deadlines, meetings, etc. and facilitating other time management tasks;
- Bringing instructors into contact with colleagues sharing the same field or interests (via their teaching agents).

Learners have their own *student agents*, which intelligently search for content and expertise; perform bookkeeping; set up (online) meetings with the instructor and other students; facilitate collaborative work among students; remind (teams of) learners of important deadlines; help them to document, publish, and archive their work; and so on (Anderson 2004: 51f.; Anderson & Whitelock 2004: 5). Finally, *content agents* are responsible for content rights management, automatic content updates, tracking learner use of content, etc. (Anderson 2004: 52). Teacher agents, student agents, and content agents are situated in the emerging Semantic Web (pp. 51ff., cf. Chapter 12.1.2.5). Another proposal in this context involves *learning support agents* that roam a web of semantically marked-up (annotated) educational resources[44] to find, retrieve, filter, sequence, and present materials from different resources according to the abilities, needs, and preferences of the individual learner (Devedzic 2004).

Learner-context interaction is another type of educational interaction proposed in the literature (Gibson 1998b). It contributes the view that learners are simultaneously active in multiple contexts extending beyond the learning environment and including family, workplace, peer groups, government, mass media, organized religion, and other sociocultural contexts (Woods & Baker 2004).

Still other interaction types consider interactions taking place within a single entity involved in the educational process rather than between several entities. *Learner-self interactions*

[44] *Educational resources* are units of specialized learning content of various kinds and sizes, including video clips, animations, simulations, Virtual Sessions (cf. Chapter 2.4.1), and complete e-learning courses. Semantic markup of an educational resource may describe aspects including title, domain, author(s), central topics, learning objectives, instructional method, curriculum integration, and so on (Devedzic 2004: 58).

(Hirumi 2002) take place within each learner.[45] They include cognitive operations that facilitate processes of learning and metacognition, enabling learner self-monitoring and self-regulation (Anderson 2004: 20).[46] Learner-self interactions form Level I in a three-level hierarchical "framework for analyzing, designing, and sequencing planned e-learning interactions" (cf. Figure 25) described by Hirumi (2002). Level II of this framework comprises interactions that occur between the learner and human (learner-instructor, learner-learner, and learner-other) and non-human (learner-content, learner-interface, and learner-environment) resources (Anderson 2004: 20). Level III is a meta-level responsible for organizing Levels I and II (Godwin 2005). At this level, learner-instruction interactions inform the design and sequencing of Level II interactions with the goal of stimulating Level I interactions (Hirumi 2002: 143).

5.4.3 The Nature of Instructional Interactivity

Moore's original threefold model and the frameworks extending it all exhibit a focus on *who* is interacting and do not provide a clear definition of *what* the nature of interaction is (Woods & Baker 2004). In contrast, Wagner (1994) offered a broad definition of interaction, which aims to capture the essence of the concept while avoiding a restriction to specific types of interaction (Anderson 2003a: 130):

> *[I]nteractions are reciprocal events that require at least two objects and two actions. Interactions occur when these objects and events mutually influence one another. (Wagner 1994: 8).*

Wagner's definition moves the focus away from the entities that are interacting and towards the nature of what is happening between them. Two key elements of interactions appear in the definition: reciprocity and mutual influence (cf. Section 5.1). In distance education, these interactions are *interpersonal*, i.e. they take place between human participants in an educational context (Woods & Baker 2004). Wagner drew a distinction between this kind of human-human interaction and interactivity, which she regarded as an attribute of technology:

> *Interactivity may eventually be viewed as a machine attribute, while interaction may be perceived as an outcome of using interactive instructional delivery systems. (Wagner 1994: 26).*

Yacci (2000) provided "a structural definition of interactivity for distance education and intelligent CBT." This definition locates interactivity in the exchange of mutually coherent messages (i.e. the communication setting, cf. Section 5.3.2) but also stresses the importance of learners' perception of interactivity (cf. Section 5.3.9), as well as positive cognitive and affective effects that can occur through interactivity. The definition identifies four major attributes of instructional interactivity:

- *Interactivity is a message loop.*
- *Instructional interactivity occurs from the learner's point of view and does not occur until a message loop* ***from*** *and* ***back*** *to the student has been completed.*

---

45 Other types of self-interaction include instructor-self interaction and content-self interaction (Godwin 2005).

46 Hirumi's notion of learner-self interactions is close to Moore's original idea of learner-content interaction (Godwin 2005), which he described as "the process of intellectually interacting with content that results in changes in the learner's understanding, the learner's perspective, or the cognitive structures of the learner's mind" (Moore 1989).

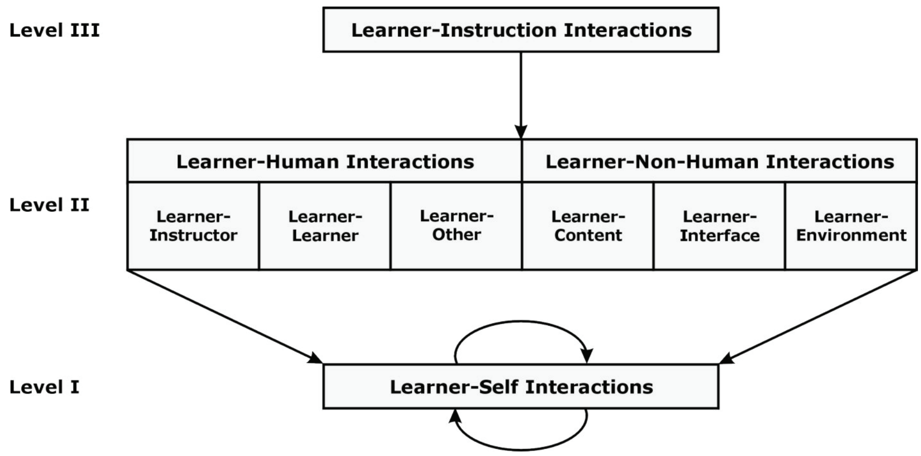

Figure 25. Three levels of planned e-learning interactions (Hirumi 2002: 143).

- *Instructional interactivity has two distinct classes of outputs: content learning and affective benefits.*
- *Messages in an interaction must be mutually coherent. (Yacci 2000: 6, his emphasis).*

Yacci's structural model defines instructional interactivity as a loop involving two entities (learners, instructors, computers and other media, pedagogical software agents, etc.) with the ability to send and receive messages. As Figure 26 shows, in this loop, messages flow from the sending entity to the receiving entity and back to the sending entity (Yacci 2000; Yacci & Hyman 2001). Examples of interactive loops include the following:

- Instructor to Learner to Instructor;
- Learner 1 to Learner 2 to Learner 1;
- Learner to Computer Program to Learner;
- Learner to Pedagogical Agent to Learner.

It is important to note that an interactive loop does not exist unless there is a message flow from the originating entity to the target entity and from the target entity back to the originating entity (i.e. not until the circuit has been closed). Messages in the structural model may involve language as well as physical actions that can be observed by both participating entities (Yacci 2000).

The second major element of Yacci's definition is the requirement that "interactivity in instruction must occur from the student's point of view" (Yacci 2000). Considering the learner's perspective in addition to that of the instructor or the computer program has important implications for the design of interactive educational experiences. First, an interactive loop may seem complete from the point of view of the instructor or the program but remains open (incomplete) from the learner's perspective because feedback from the instructor or program concerning his or her response is not provided (Yacci 2000, cf. Section 5.3.1). This view is consistent with Rafaeli's third-order dependency requirement: an exchange is fully interactive only after the third message, which relates both to the initial message and its response (cf. Section 5.3.2). The idea of feedback indicating the completion of an interactive loop has been mentioned by several authors (Weller 1988; Berge 1999; Liaw & Huang 2000; Northrup 2001). Learners expect and rely on the instructor (or program) to close the communications loop:

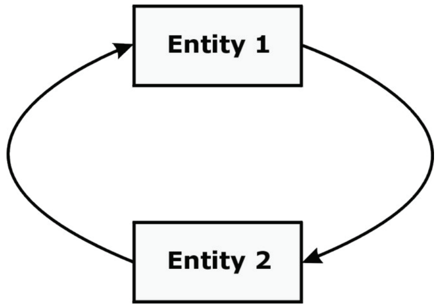

Figure 26. A completed message loop between two entities (Yacci 2000, Figure 1).

> *Until students receive a 'reply' in some form verifying that what they sent was accurate, they typically are uncomfortable. Additionally, students may not be conceptualizing concepts in the manner intended. (Northrup 2001).*

Therefore, both instructors and designers of software for learning need to make sure that they provide feedback which completes the learner's interactive loop. Second, even when feedback is given by the instructor or system, the learner may not recognize the message as related to his or her contribution and hence not perceive the response as interactive (Yacci 2000). Hence, feedback should be appropriately contextualized and delivered in a way that learners can understand. Effective use of feedback is helpful in setting up a dialogue which involves the learner as an active participant rather than as a consumer of content (Brown & Voltz 2005).

Yacci (2000) identified two distinct outcomes of instructional interactivity. The first, *content learning*, is a cognitive learning process for achieving an instructional goal, which involves the creation, extension, and modification of knowledge structures. The second outcome, *affective benefits*, includes the learner's emotions and values attached to people, content, institutions, programs, etc. involved in instruction, which may be intensified or reduced in the instructional process.

Finally, in Yacci's model, *mutual coherence* between a message and its response refers to the extent that the content ('payload') of the message sent and the content of the responding message are related to each other (refer to the same topic). Yacci argued that "the extent of shared meaning between messages influences the perceived degree of interactivity" (Yacci 2000). A high degree of shared meaning between messages leads to high mutual coherence and thus contributes to learner impressions of high interactivity (Yacci 2000). It should be noted that mutual coherence of messages does not consider how relevant an exchange is to instructional goals. The content of the exchange may even be irrelevant from a content learning perspective but may produce significant affective benefits (Yacci 2000), such as aiding in relationship building (cf. Section 5.4.5 and Chapter 7.16).

Yacci et al. (Yacci 2000; Yacci & Hyman 2001) discussed several variables that influence the quality of the interactive loop, including lag time of response, message duration, amount of information content, and multiplexing. *Lag time of response* refers to the delay between sending a message and receiving a response that closes the loop. In synchronous communication, responses are generally expected within a few seconds. In contrast, the time lag between messages in asynchronous communication (mail, electronic mail, discussion forums, etc.) can be longer, but not too long. Again, it is important to include the learner's point of view. If a learner sends an e-mail to his or her instructor and the instructor responds only five weeks later, although within minutes after reading it, there is no significant delay in response from the instructor's perspective, but the perceived response lag for the learner is high (Yacci 2000).

Northrup et al. (2002) studied learner perceptions of online interaction. They found that "[p]participants strongly stated that the need for timely responses from peers and from their instructor was of utmost importance" (p. 6). In general, the learners of today have grown or are growing up in a world in which cell phones, MP3-players, computer games, the World Wide Web, e-mail, instant messaging, and other digital technologies are natural elements of their daily lives. These 'new' learners, called the "Net Generation" (Oblinger & Oblinger 2005) or the "Digital Natives" (Prensky 2001a; Prensky 2001b; Woodill 2004), as opposed to the "Digital Immigrants" generation of their parents and instructors, who were born into a non-digital world and entered the digital realm only as adults, "crave *interactivity* – an *immediate* response to their each and every action" (Prensky 2001b: 4, emphasis added). An educational system teaching the Net Generation with essentially the same approaches and technologies as it used on their parents is generally not well equipped to meet this demand. Digital Natives have low tolerance for response lags and quickly lose interest if these lags become significant, a fact that needs to be considered when designing online courses and other learning experiences for the Net Generation.

Concerning the response times of interactive loops between learners and computers, Tognazzini (1993) argued with respect to software in general that responses which the machine can give at once should not be delayed. When the system needs some time to respond, learners should be reassured with immediate intermediary feedback, followed by displays, activities, etc. that keep them "interested and entertained" (in other words, occupied) during the waiting period in order to reduce their perception of the response lag (Tognazzini 1993).

Message duration is another variable that refers to the length of time expired between the beginning and the end of a message (Yacci 2000; Yacci & Hyman 2001). The duration of messages in interactive exchanges is generally expected to be short or moderate, otherwise the interaction might be perceived as consisting of monologues; however, the tolerable length of a message may differ from one situation to the next (Yacci 2000). Yacci continued to point out that while message length has an impact on the mutual coherence of messages, the *amount of information content*, i.e. the quantity of new information included in a message, independent of its length, also contributes to mutual coherence. If the content of a message can be largely predicted by its recipient, the message is said to exhibit a high degree of *information redundancy* and a low degree of *uncertainty* or *entropy* (Shannon 1948a; Shannon 1948b) because it contains little information that is not known to the recipient. "The more we know about what message the source will produce, the less uncertainty, the less the entropy, and the less the information" (Pierce 1961: 23). In instruction, this is true of confirmatory feedback messages to *correct* learner responses. Yacci concluded that feedback to correct responses should be kept concise because it does not provide any new information. He went on to argue that feedback can be more effective if it follows *incorrect* responses because then the feedback message contains more information unknown to the learner (lower information redundancy; higher uncertainty and entropy) and may therefore be perceived as having higher information content, contributing to perceptions of higher degrees of interactivity. Redundant information in instructional responses is not always negative, though. In fact, it can help to reassure learners who lack confidence while they are doing some task. In these cases, information redundancy has benefits relating to learner affect rather than content learning (Yacci 2000).

Concerning the nature of the content of instructional messages, Yacci (2000) pointed out that messages in an educational context may combine (*multiplex*) both cognitive and affective information into an aggregate contribution, which, for example, provides the necessary scaffolding (cf. Chapter 12.3.2) and encouragement (cf. Chapter 12.3.8) to help the learner to accomplish some task. In face-to-face communication, humans (and increasingly embodied interactive software agents) naturally convey the affective component of a message by means of performing the various non-verbal communication behaviors discussed in Chapter 10.1 (speech, facial expression, gaze, gesture, posture, and locomotion), which Yacci (2000) referred to as *paralanguage*. In contrast, communication of affect is not a natural element of interactions in computer-assisted instruction and e-learning, especially as far as automated responses are concerned, but has to be explicitly designed into the educational experience. This is important because learners might interpret the lack of perceptible affective cues in instructional messages as the indicator of an indifferent or uncaring attitude of the instructor or system (Yacci 2000). Embodied pedagogical software agents can be designed to combine affective and cognitive information into multimodal messages which are delivered using verbal and non-verbal means.

As in interactive experiences in general, the *perceived interactivity* (cf. Section 5.3.9) of educational processes and contexts influences learners' perception of their own performance and their satisfaction with the educational experience. Swan (2001; 2002) conducted a large survey-based study of learners in online courses to elicit how they perceived their courses along various dimensions, including learning, interaction with the instructor, interaction with other course participants, their personal level of activity, and their satisfaction with the course. An important result of her study was that learners who *believed* that they had more interactions with both their instructors and peers reported higher levels of satisfaction with their courses and thought that they learned more (Swan 2001). Furthermore, Swan found significant positive correlations between learners' reported levels of activity in the course and their reported satisfaction and learning, as well as between greater structural consistency within the course and learners' satisfaction and perceived learning. Hence, "frequent and engaging interaction with course content" and "a clear and consistent course structure" also contribute to learners' satisfaction and perceptions of learning (Swan 2002). In summary, Swan's study showed that from learners' point of view, interaction with content, instructors, and other participants plays an important role in successful online learning and therefore should be considered by instructors and developers of online courses alike.

Other researchers came to similar conclusions. Flottemesch (2000) reviewed the literature on interaction in distance education and found support for the hypothesis that learners' judgment of distance courses depends on their perception of their interaction with the instructor. This view is shared by a number of other authors cited by Woods and Baker (2004). According to these authors, learners' perception that the level of interaction with instructors and other learners is 'sufficient' has a positive impact on their satisfaction with the learning experience because it helps them to develop a "sense of personalization and customization of learning" (Boettcher 1999: 43) and to get over feelings of distance and isolation (Woods & Baker 2004). If the interaction is perceived to be difficult, the level of learner satisfaction declines, whereas satisfaction increases when learners perceive that the course instructor places importance on interaction (Arbaugh 2000: 44ff.).

5.4.4 Transactional Distance

The "universe of teacher-learner relationships that exist when learners and instructors are separated by space and/or time" (Moore 1993: 22) is the subject of Moore's theory of *transactional distance* (Moore 1972; Moore 1991; Moore 1993), which provides an account of the relationships between the parties involved in a distance learning setting (Bodomo et al. 2001). Moore (1993) stated that the separation of learners and instructors in distance education results in certain behavioral patterns of the participants and influences both teaching and learning processes. According to Moore, learners and instructors are separated by a "psychological and communications space" with an inherent potential for misunderstanding between learner and instructor inputs (Moore 1993). This space, called the *transactional distance*, has to be crossed in a distance education setting. Moore defined transactional distance as a function of the dialogue (interaction) between learners and instructors, the structure of the course design, and the nature and degree of learner autonomy (Moore 1993). As summarized by Wallace (2003), Moore's conjecture was that a smaller transactional distance indicates greater learner involvement. More dialogue (i.e. purposeful, constructive, and valued interaction(s) (Moore 1993)) between learner and instructor would be a sign for a smaller transactional distance, whereas more structure provided by the instructor would result in a larger transactional distance (Wallace 2003: 245). As the role of dialogue becomes more prominent, the structure imposed by the course design, including instructional goals, activities, and assessment, decreases to allow for more learner control (Bodomo et al. 2001). In contrast, more structured educational experiences reduce students' control of their own learning, which, in turn, diminishes their engagement with the content (Wallace 2003: 245).[47] An example of high structure and low dialogue (and hence a large transactional distance) is the traditional lecture with no room for questions or discussion, whereas synchronous online discussions among learners exhibit a smaller transactional distance (low structure and high dialogue) (p. 245). Learner autonomy is a third factor which interacts with dialogue and structure and is defined as the extent to which the learner rather than the teacher determines the goals, experiences, and evaluation of his or her own learning (Moore 1993). Moore described the relationship between dialogue, structure, and learner autonomy as follows: more structure and less dialogue (i.e. greater transactional distance) in a distance learning scenario require the learner to exercise more autonomy, by taking responsibility, making judgments, and taking decisions with respect to his or her studies (Moore 1993).

Pedagogical software agents can help to reduce transactional distance even in highly structured online learning scenarios with few possibilities for dialogue between learners and human instructors, by involving learners in instructional conversations with virtual coaches, experts, guides, etc. that are capable of various kinds of interactions available to human instructors (cf. Chapter 12.3). With their ability to monitor and adapt to the individual learner, these agents furthermore allow the learning environment to accommodate and be responsive to individual learners' needs (cf. Chapter 4.2).

5.4.5 Social Interaction

Interpersonal interactions between and among learners and instructors are commonly not only about the instructional content or the task at hand (the *educational* dimension) but also involve a *social* dimension that has less do to with topics of the course and more with building trust, relationships (cf. Chapter 7.16), and community among the participants (Kreijns et al. 2003: 342, cf. Figure 27). The term *social interaction* refers to exchanges between two or more individuals which aim to "transcend psychological distances and establish interpersonal

[47] In contrast, Borsook and Higginbotham-Wheat (1991) argued that the content and/or the instructor can provide structure which offers learners guidance and scaffolding. Leaving complete control to the learner has the inherent danger of causing confusion and loss of orientation (Godwin 2005).

connections" (Wagner 2006: 48) between them in order to facilitate both cognitive and socio-emotional processes involved in learning.[48]

Superficially, it could be argued that social interaction serves little purpose with respect to achieving instructional goals and poses an unnecessary source of distraction for learners, but, in fact, social interaction has been widely recognized as a key element of educational experiences in general and online courses in particular (e.g. Wagner 1994; Gilbert & Moore 1998; Liaw & Huang 2000; Northrup 2001; Kreijns et al. 2003). Students in online courses appreciate and benefit from interactions with their peers and the instructor (Wallace 2003: 250). When there is more social rapport among the participants in an online course, learners are said to experience less psychological distance, isolation, and frustration (Wolcott 1996). In addition, social interaction not only increases the total amount of interaction of an online course (Roblyer & Wiencke 2004: 27) but also directly fosters instructional interaction (Gilbert & Moore 1998: 31) and thus enhances and furthers learning (Mason & Rennie 2006: 62). Many researchers in (distance) education share the view that social interaction is essential for effective (online) collaboration and communication and inseparable from it: "If there is collaboration then social interaction can be found in it, and vice versa, if there is no social interaction then there is also no real collaboration" (Kreijns et al. 2003: 338). Before they can effectively collaborate with their peers, learners need to build rapport with them in the early stages of forming a learning group, for which social exchanges are deemed a safer vehicle than content-related discussions (Mason & Rennie 2006: 63). This point was also emphasized by the following author:

> *[I]f students are to offer their tentative ideas to their peers, if they are to critique the ideas of their peers, and if they are to interpret others' critiques as valuable rather than as personal affronts, certain conditions must exist. (...) [S]tudents need to trust each other, feel a sense of warmth and belonging, and feel close to each other before they will engage willfully in collaboration and recognize the collaboration as a valuable experience. (Rourke 2000).*

Figure 27 summarizes the dual function of social interaction in learning experiences. On the one hand, social interaction facilitates socializing, getting to know each other, committing to social relationships, developing trust and belonging, building a sense of community, and other socio-emotional processes involved in relationship building, group formation, and group dynamics (Kreijns et al. 2003: 342). On the other hand, social interaction is also a central component of individual and collaborative cognitive processes of learning. *Social development theory* (Vygotsky 1962; Vygotsky et al. 1978) emphasizes the fundamental role of social interaction in the development of cognition (Bodomo et al. 2001). Social interaction is the basis for critical thinking, helps to develop shared understanding within a group of learners, enables the social construction of knowledge, and supports the acquisition of competences (Kirschner et al. 2003: 323, cf. Chapter 2.2.3).

Social interaction is the preferred means for individuals to cross the psychological and communicative gap separating them from other participants in an educational experience (cf. Section 5.4.4) and to establish a sense of connectedness with their peers and instructors (Wagner 2006: 48). The degree to which an individual feels real ('present'), or is perceived that way by other participants, in an online learning environment positively influences his or her perception of the strength of his or her connection with them, leading to greater overall satisfaction with the experience and greater willingness to stay within the environment or

[48] It should be noted that the term "social interaction" is used in this section in a narrower sense than in the rest of the thesis. Here, the focus is on the interpersonal dimension and on relationship building (cf. Chapter 7.16).

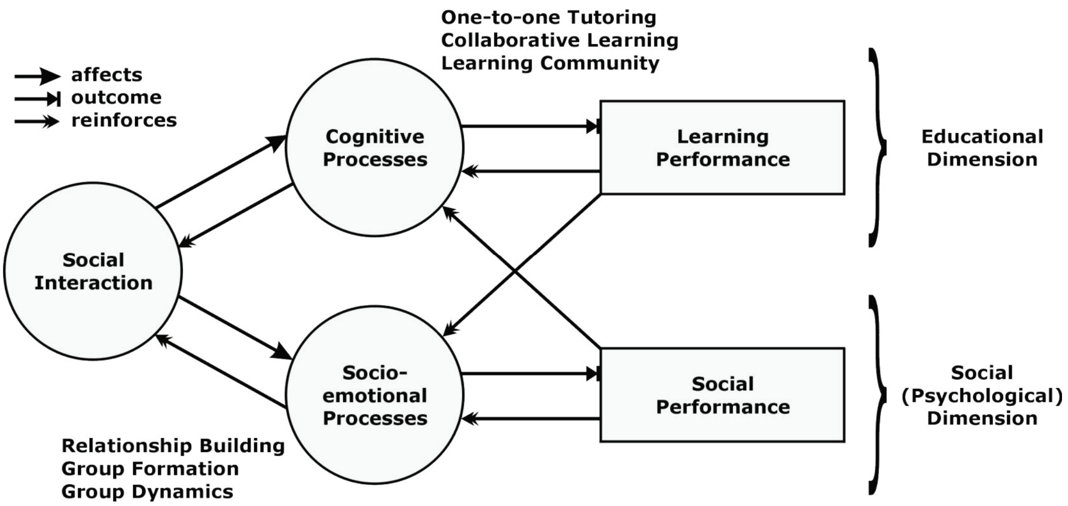

Figure 27. The two functions of social interaction in learning. Adapted from Kreijns et al. (2003: 342, Figure 1).

experience (p. 48). The notion of *social presence* has been used to describe this perception. In general terms, *presence* is "the perceptional illusion of non-mediation" (Lombard et al. 2000: 77) created by an environment mediated by technology. Social presence is an important dimension of presence, where people feel that other intelligent entities co-exist and interact with them in the environment (Heeter 1992). According to Biocca, social co-presence may be indicated by a form, behavior, or sensory experience. Its amount depends on the degree to which an individual has a sense of access to the intelligence, intentions, and sensory impressions of another entity (Biocca 1997). In the same vein, Lee defined social presence as "a psychological state in which virtual (para-authentic or artificial) actors are experienced as actual social actors in either sensory or non-sensory ways" (Lee 2004: 45). In this state, technology users fail to detect both the para-authenticity of mediated humans and the artificiality of non-human agents (p. 45; Lee et al. 2006: 964).

Social presence theory (Short et al. 1976) is concerned with how different media influence perceptions of interpersonal communication regarding satisfaction and efficiency (Dirkin et al. 2005: 117). This theory views social presence as a subjective quality of the medium used to communicate with other people (Short et al. 1976: 65). The social impact of different communication media is argued to depend on the degree of acoustic, visual, and physical contact between communicators which these media permit ([INT 43]). Social presence is defined as "the degree of salience of another person in an interaction and the consequent salience of an interpersonal relationship" (Short et al. 1976: 65). Key components of social presence include intimacy, immediacy, warmth, and interpersonal rapport ([INT 43]), which can be increased by providing more channels that can transmit relevant cues (such as facial expressions, body movement, and eye contact) between communicators (Lombard & Ditton 1997). Different media vary in their capacity to convey verbal and non-verbal cues that facilitate the perception of other participants in the communication as physically salient and present (Kreijns et al. 2003: 345).[49]

[49] One implication of social presence theory is that media which are less capable of transmitting non-verbal cues, in particular media that are text-based, such as e-mail and instant messaging, do not allow higher degrees of social presence because of the lack of social and affective information, resulting in negative effects on interpersonal communication (Rourke et al. 1999; [INT 43]). This

Media providing a larger number of channels for communicating verbal and non-verbal cues can enhance users' sense of social presence (Lombard & Ditton 1997) because they allow users to obtain richer information about the degree of realness of others as well as their own within the environment. The range of entities which are perceived as real, or having a social presence, is not limited to people. Lombard and Ditton noted that one type of presence is that of "social actor within medium," which arises from perceptions and psychological processes that cause people to "illogically overlook the mediated or even artificial nature of an entity within a medium and attempt to interact with it" (Lombard & Ditton 1997). Pedagogical agents capable of embodied conversation (cf. Chapter 11.3.4) and endowed with expertise (cf. Chapter 12), emotional intelligence (cf. Chapter 13.4), and personality (cf. Chapter 13.5) are such social actors within the medium that invite learners to suspend their disbelief (cf. Chapter 3.1.3). They can serve as a substitute for the social presence of instructors and other learners in computer-based learning environments, creating a better sense of human-like interaction than other forms of learning from computers by broadening the bandwidth of social communication between computers and learners and increasing learner engagement and motivation (Johnson et al. 2000). Early empirical investigation (Lester et al. 1997b; Lester et al. 1997c) revealed the *persona effect*, which is that the mere presence of a pedagogical software agent, even if the agent is not expressive, "can have a strong positive effect on student's [sic!] perception of their learning experience" (Lester et al. 1997b). Later research found some support for the persona effect but also cautioned against believing in its unconditional validity (van Mulken et al. 1998; Dehn & van Mulken 2000; Moundridou & Virvou 2002; Dirkin et al. 2005; Yee et al. 2007). For example, the participants in the study conducted by Dirkin et al. (2005) perceived more social presence both when a computer-based tutorial only consisted of text and graphics and when it additionally featured an animated presentation agent using locomotion and pointing gestures and reading the text of the tutorial in a modulated voice with synchronized lip movements, adding commentary now and then to increase its believability as a social entity. Furthermore, participants in the text-only condition perceived the learning experience more positively. Given the 'paradoxical' results of their study, Dirkin et al. emphasized the need for

> *careful design of social behaviors for animated pedagogical agents if they are to be of educational value, otherwise, the use of agent technology can actually detract from the learning experience. (...) [I]n certain contexts it may actually be better to have no agent than a badly designed one. In other words, as in most things in life,* ***either do it well or don't do it at all.*** *(Dirkin et al. 2005: 113 + 125, emphasis added).*

A number of researchers (e.g. Berge 1999; Northrup 2001; Kreijns et al. 2003) have stressed that neither social nor task-/content-related interaction happens automatically but has to be explicitly designed into an online course (or any other educational experience). Therefore, designers and instructors of online courses commonly create activities and technologies intended to facilitate communication and collaboration and to promote social presence and community (Mason & Rennie 2006: 63). Given the positive connection between interaction and students' (perceived) learning and satisfaction (cf. Section 5.4.3), it is tempting to incorporate more and more interaction into a course. However, an overuse or misuse of interaction may

---

suggests, in turn, that social and affective interaction cannot be adequately supported by computer-mediated communication, which is still largely text-based (Rourke et al. 1999, cf. Chapter 2.3.8), a phenomenon which has been termed the *computer-mediated communication paradox* (Hudlicka 2003: 22). It should be noted, though, that this view has been contested by later reviewers (Rourke et al. 1999) and empirical studies (e.g. Burgoon et al. 2002).

backfire on the designer or instructor since it can lead to learner experiences of boredom, frustration, and overload (Berge 1999; Northrup 2001; Northrup et al. 2002). On the other hand, too little interaction is likely to increase learners' psychological distance, isolation, and frustration and hence decrease learning and satisfaction (Northrup et al. 2002).

5.4.6 Interactivity in Learning from Computers

Research on distance education nowadays tends to focus on interpersonal communication with instructors and other learners *through* computers (Bannan-Ritland 2002; Kreijns et al. 2003) as well as learning *with* computers as cognitive artifacts (cf. Chapter 6.4) or 'mindtools' (Lajoie & Derry 1993; Jonassen & Reeves 1996; Jonassen et al. 1998; Lajoie 2000), subscribing to current constructivist views of learning, sometimes at the expense of considering processes of learning *from* computers through learners' interaction with content presented in a computer interface, which are often regarded as obsolete (Yacci 2000, cf. Chapter 2.2.4). In contrast, Sims (1997a) explicitly considered the human-computer dimension, in particular "what makes an application interactive, instructional, and effective" (Sims 1997a: 158). In a later paper, he defined interactivity from a human-computer perspective as follows:

> *Interactivity refers to the facilities provided by a computer-based application to provide the user with both control of the process and communication with content. This communication involves both the user initiating an action and the computer responding to that action. The computer's role with interactivity is to enable the processing power to interpret and respond to a user's action. (Sims 1999).*

Taking the point of view of a multimedia courseware developer, Sims (1997a) proposed a framework that can be applied to the analysis of learner-content interaction in computer-based learning environments. Similar to Yacci's (2000) notion of the interactive loop (cf. Section 5.4.3), Sims (1997: 159f.) discussed the necessity of setting up an "interaction circuit" between learner and computer, where the learner participates in a "true dialogue" with the computer-based application. "It is this dialog [sic!] or circuit which we must pursue, enabling the users to be continually and productively active while working with the instructional content" (Sims 1997a: 160). This view of interactivity goes beyond simple selection of menu items, clickable elements, linear sequencing of content, and the other elements of what has often been labeled 'interactive' educational software (pp. 157f., cf. Chapter 2.3), moving towards interaction that is "involving and personal" (Dickinson 1995: 145). The classification proposed by Sims consists of a set of interactive concepts which he intended to serve as a guide to the different options for interactivity available to developers of multimedia courseware (Sims 1997a: 162–168). Sims noted that these interactive concepts are not mutually exclusive but can be combined to create more sophisticated and engaging learning experiences (p. 162), as in the last two types (non-immersive contextual and immersive virtual interactivity).

- *Object interactivity*. The user activates (clicks on) objects (buttons, people, and things) using some kind of pointing device. The activation produces an audio-visual response.
- *Linear interactivity*. The instructional content is pre-organized as a linear sequence through which the learner can navigate forwards and backwards ("electronic page-turning"), without receiving response-specific feedback to his or her actions (cf. Chapter 2.3.2).
- *Hierarchical interactivity*. The application provides menu-based navigation, offering several predefined options from which learners may choose a specific instructional path. In other words, learners are given more choices than in linear interactivity.

- *Support interactivity*. The learner has access to different forms of performance support, i.e. "an electronic system that provides integrated access to information, advice, learning experiences, and tools to help someone perform a task with the minimum of support by other people" (Raybould 1995: 10). Performance support ranges in sophistication from help messages through FAQ (frequently asked questions) collections and knowledge bases to complex tutor, guide, and help desk agents (Franke 2006: 284–290). In distance learning, support should be available anytime and anyplace (Northrup 2001).[50]
- *Update interactivity*. The learner engages in a dialogue with instructional content generated by the application. He or she is expected to respond to problems presented or created by the system (possibly according to the learner's previous performance). The system analyzes the learner's contribution and responds with appropriate updates or feedback. The complexity ranges from simple question-and-answer formats (cf. Chapter 2.3.1) to the context-sensitive and learner-adaptive responses of intelligent tutoring agents (cf. Chapter 2.3.3 and Chapter 12.3).
- *Construct interactivity*. The learner manipulates objects in the learning environment, which may involve building, re-arranging, dissecting, etc., in order to attain certain goals related to learning. This type of interactivity is closely linked to simulation interactivity because it helps to create authentic learning environments, but without the risk and costs associated with real-world situations (cf. Chapter 2.3.4).
- *Reflective interactivity*. The application has a memory of responses given by other learners or recognized experts on the subject area, to which learners can compare their own answers. Hence, the learner has the opportunity to reflect on and assess the quality of his or her response. This strategy is useful in domains where it is difficult (in principle or for the machine) to evaluate the accuracy or correctness of the user's response.
- *Simulation interactivity*. This type of interactivity is closely related to construct interactivity. The simulation places the learner in the role of controller or operator with the ability to manipulate certain variables within the system (cf. Chapter 2.3.4). The learner's actions can produce directly observable effects (updates) within the simulation; however, generation of an appropriate update may also require that the learner has completed a particular sequence of actions first.
- *Hyperlinked interactivity.* The system provides abundant information for learners to explore, which is organized as a hyperlinked structure (cf. Chapter 2.3.6). The hyperlinking enables learners to access the information flexibly by following their own paths through the link structure. All possible (or relevant) paths have to be accessible. Learners also need tools that facilitate navigation and orientation.
- *Non-immersive contextual interactivity.* The application integrates and extends the previous interactive concepts to create a complete virtual learning environment, which models a real-world work environment or other setting and engages the learner in authentic tasks (cf. Chapter 16.1).
- *Immersive virtual interactivity.* The learner is 'transported' into an immersive computer-generated world which responds to his or her movement and actions and in which the learner experiences a sense of presence (of being 'there') in the environment and co-presence (that others are present in the environment together with the learner) (cf. Section 5.4.5). For a long

[50] While costly to realize with human resources, pedagogical agents can be designed to fill various supportive roles that meet the requirement of constant availability of performance support in online courses (cf. Chapter 16.2).

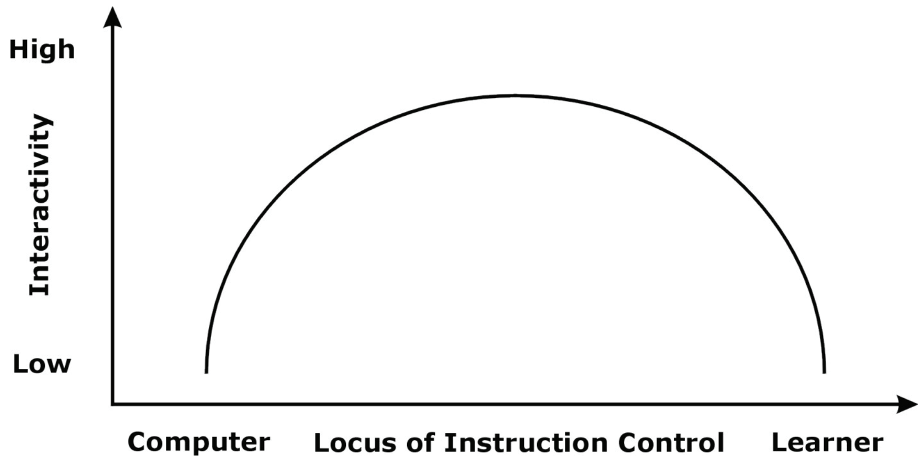

Figure 28. Interactivity and locus of control in learner-computer interactions. Adapted from Borsook and Higginbotham-Wheat (1991). According to Borsook and Higginbotham-Wheat, the level of interactivity reaches its maximum at a medium point between complete computer and complete learner control.

time, such environments were largely science fiction; however, with the recent availability, growing sophistication, and exploding popularity of synthetic worlds for gaming and socializing (Castronova 2006), such as World of Warcraft ([INT 44]) and Second Life ([INT 45]), creating immersive educational experiences has become much more than a possibility (Bell & Trueman 2008). In these worlds, users cannot only interact with virtual objects but, importantly, also with other users (represented by their avatars) as well as embodied interactive software agents with different roles and complexity (cf. Chapter 16.2).

Sims integrated his interactive concepts into a more comprehensive *engagement-control model* of interactivity, which defines three dimensions of interactive computer-based instruction (Sims 1997a: 168f.). *Engagement* is the first dimension, which refers to the nature of the user's involvement and is either navigational (the user moves through the application) or instructional (the user interacts with the content). The second dimension, *control*, is concerned with the relative degree of learner vs. program control over making decisions regarding instruction or navigation. The third dimension is represented by the *interactive concepts* discussed above and indicates the type of interaction expected in different constellations of the model.

Regarding the *locus of control* in the interaction between learner and computer, with a focus on pacing and sequencing, Borsook and Higginbotham-Wheat (1991) suggested that an optimal level of interactivity is neither associated with complete learner control nor with complete program control but involves a balance of control between learner and computer (cf. Figure 28). This view is in line with the general idea of the superiority of mixed-initiative human-machine dialogue over both system-initiative and user-initiative dialogue strategies (cf. Chapter 11.3.3.2.1). However, in real learner-computer scenarios, the locus of control may shift from the medium point either towards the learner's or the computer's end, depending on additional factors, such as the nature of the task, the content, learners' metacognitive skills and level of development, etc. (Godwin 2005).

Each of the interactive concepts proposed by Sims (1997a) represents a particular level of interactivity in the interaction of learners with computer systems. Figure 29 locates the different types of interactivity along a continuum from low to high interactivity. A similar scale for interactivity was proposed by Schulmeister (2003; 2005). Referring to screens of multimedia courseware or web pages in e-learning platforms, he developed a taxonomy of the multimedia

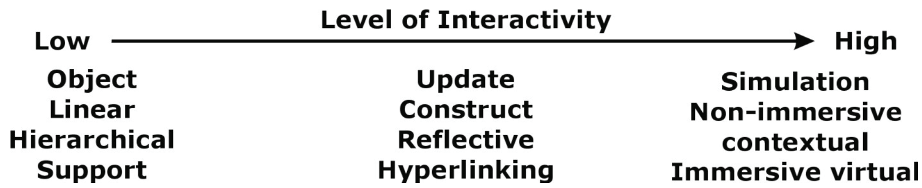

Figure 29. Sims's interactive concepts arranged according to the level of interactivity in the interaction between learner and computer system.

elements contained in a screen or web page, including text, images, diagrams, animations, audio and video clips, tables, formulas, and small embedded applications (Schulmeister 2005). The levels in this taxonomy differ with respect to the level of learner-content interactivity provided (Schulmeister 2003):

- *Level I*. Viewing objects and receiving information;
- *Level II*. Watching and receiving multiple representations;
- *Level III*. Varying the form of the representation;
- *Level IV*. Manipulating the content of multimedia representations;
- *Level V*. Constructing the object or representation contents;
- *Level VI*. Constructing the object of the representation and receiving intelligent feedback from the system through manipulative action.

Schulmeister (2003) argued that even at the highest level of interactivity, current learning systems have not reached the human capacity for communication and social interaction yet (see also Sims (1997b)). However, equipped with increasingly sophisticated verbal and non-verbal behaviors (cf. Chapter 10.1), language processing and conversation capabilities (cf. Chapter 11), expertise (cf. Chapter 12), and emotional competence (cf. Chapter 13), pedagogical agents can cross the gap between 'artificial' human-computer dialogue and 'natural' interpersonal communication and participate in instructional exchanges with the learner (cf. Chapter 12.3) that involve meaningful actions and conversation. With their capacity to make the computer an intelligent conversational partner in learning, pedagogical agents can improve the humanness and adaptiveness of learners' interactions with computers and thus improve the ability of people to learn from them (Yacci 2000).

5.5 Discussion and Summary

While interactivity has gained currency in the popular and scientific literature and media outlets and people may have an intuitive grasp of the concept ("I know it when I see it"), the exact nature of interactivity remains controversial. Bucy admitted that "[a]fter nearly three decades of study and analysis, we scarcely know what interactivity *is*, let alone what it *does*, and have scant insight into the conditions in which interactive processes are likely to be consequential for members of a social system" (Bucy 2004b: 373, his emphasis) and that "we are still far from a coherent *theory* of interactivity, with an agreed-on set of principles and core assumptions" (Bucy 2004a: 371, his emphasis). Nevertheless, it is possible to extract from the various conceptualizations of interactivity reviewed in Section 5.3 a number of aspects that are essential ingredients of interactive experiences and thus instructive for the designers of embodied interactive software agents:

- Interactivity involves *feedback*: the recipient of a message can send a response which influences the sender's next message (cf. Section 5.3.1).
- Interactivity is not an absolute (yes-or-no) concept but a matter of *degree* (cf. Section 5.3.2), ranging from low to high levels of interactivity.
- Interactive communication exhibits *third-order dependency* among messages, i.e. the third message in an exchange is a reaction to the response sent to the first message, relating both to the response and the initial message (cf. Section 5.3.2).
- Interactive communication is a *two-way (or multi-way) process* in which the roles of sender and recipient of messages are *interchangeable* (cf. Section 5.3.3).
- The participants in an interactive exchange, including both humans and machines, are more or less readily *responding* to each others' contributions or communication needs (Ha & James 1998: 461), ideally with no perceptible delay (cf. Section 5.3.4).
- Interactivity involves giving users some degree of *control* over the content, form, and pace of a mediated environment populated by humans and/or machines (cf. Section 5.3.5).
- There is an emphasis on short response times in interactive exchanges. Preferably, the exchange happens in *real time* while users can still control the rate of the exchange (cf. Section 5.3.6).
- Interactivity as a concept may include *non-mediated* face-to-face communication or may be restricted to *mediated* interaction (cf. Section 5.3.7).
- *Interactivity-as-process* refers to the interaction between humans mediated by technology, whereas *interactivity-as-product* is concerned with interaction processes between humans and technology (cf. Section 5.3.8).
- Interactivity is increasingly regarded not as an attribute of the technology (medium) or the communication setting but as something that resides, at least partially, in the *perceptions* of individuals (McMillan 2002). Interactivity as user perception emphasizes the subjective, experiential dimension of the concept (cf. Section 5.3.9).
- Interactivity is a *multidimensional* construct. Each dimension is represented by a continuum (cf. Section 5.3.10). The multidimensional space has been defined and partitioned in different ways that emphasize different views on the nature of interactivity.

Recently, the focus of the debate on interactivity has shifted from concerns about definition and descriptive typology to research of a more systematic and cumulative nature (Bucy 2004a: 371). The emerging view in analyses is that "interactivity is not a fixed concept, associated with a particular information technology or mode of communication, but may reside in various locations, whether as a feature of media systems, message exchanges, or user perceptions" (Bucy 2004a: 371). Concerning its value in interactions between humans or between humans and machines, Burgoon et al. pointed out that

> *interactivity is neither inherently positive nor negative. The extent to which interactivity helps or hinders is a function of a variety of factors including the nature of the interaction, the objectives to be accomplished, and the interactants themselves. (Burgoon et al. 2002: 660).*

It has been argued that interactive experiences should be classified as either 'process' (human interaction with other humans through technology) or 'product' (human interaction with technology) (Stromer-Galley 2004, cf. Section 5.3.8). For as long as human-computer interactions have been unable to reach the flexibility and sensory richness of (face-to-face) communication between humans, the classification of interactive experiences with respect to this dichotomy has been unproblematic. But as embodied software agents are becoming increasingly capable of acting as socially competent interaction partners that can use the same verbal and non-verbal cues which are available to people in face-to-face situations, the line between human-human and human-computer interaction is blurred. Embodied interactive software agents (cf. Chapter 3.3) are both: social actor and technology. As their sophistication grows, human interaction with them will move closer to the human end of the continuum. Since interactivity is always also in the eye of the beholder (cf. Section 5.3.9), in users' perceptions, agents might even reach human levels of interactivity before their capabilities actually allow them to get there.

In (distance) education, interaction and interactivity are regarded by most researchers and practitioners as central elements of educational processes, both face-to-face and at a distance (cf. Section 5.4). As discussed in Section 5.4.1, learning experiences that provide more (sophisticated) interaction (or interactivity), in particular if they involve learning from, with, and through information and communication technologies, are commonly expected to result in improved learning outcomes and to increase learner satisfaction. Furthermore, they are highly compatible with constructivist views of learning (cf. Chapter 2.2.3). While few seem to doubt the benefits of interaction and interactivity in education, a definition of either term that appeals to everyone (or at least to most) is yet to appear (as for interaction and interactivity in general). One approach to capture the elusive concept of interaction in education identifies different types of interaction (cf. Section 5.4.2), including learners' interaction with the content, the instructor, other learners, the interface, the environment, and the sociocultural context, as well as interaction between and among instructors and content. Anderson's equivalency theorem implies that learner-content, learner-instructor, and learner-learner interaction are equivalent in that substituting one type for one of the others does not degrade the educational experience (Anderson 2003b: 4). An interesting corollary of the equivalency theorem is the option to automate certain interactions by employing teacher, student, and content agents. Apart from interactions *between* entities in the learning process, there are also interactions *inside* an entity, such as learner self-interactions.

Shifting the focus away from who is interacting, the nature of instructional interactivity was discussed next in Section 5.4.3, starting with Wagner's (1994) broad definition of interaction, whose key elements include reciprocity and mutual influence. The structural definition by Yacci (2000) describes interactivity as a complete message loop from and back to the learner, which involves the exchange of mutually coherent messages and produces cognitive and affective benefits. One important element of Yacci's conceptualization is the emphasis on the learner's point of view. In general, the perceived interactivity of educational processes has been found to have a positive influence on learners' perception of their own performance and their satisfaction with the learning experience, for example in Swan's study (Swan 2002).

Moore (1972; 1991; 1993) identified a "psychological and communications gap" (Moore 1991: 2f.) between the participants in a distance learning setting, the "transactional distance," which he saw as a function of the dialogue (interaction) between learners and instructors, the structure of the course design, and the nature and degree of learner autonomy. It was suggested in Section 5.4.4 that transactional distance in online learning settings with much structure and little dialogue between the human participants could be reduced with the help of pedagogical agents engaging the learner in various kinds of instructional interactions.

Interpersonal interactions in education are not only concerned with the instructional content or task but also involve social interaction among the participants to establish connections, relationships, and community (cf. Section 5.4.5). Social presence, i.e. the degree to which learners feel real ('present') or are perceived that way by others in an online learning setting, positively influences the perceived strength of interpersonal connections and thus contributes to learners' satisfaction and loyalty. Social presence theory (Short et al. 1976) implies that media providing more channels for contact between communicators can increase users' sense of social presence through their ability to transmit richer information about their own perceived realness and that of others. Not only people are perceived as real, or having a social presence, but also artificial entities like agents. Embodied pedagogical agents are social actors within the medium that can serve as a substitute for the social presence of instructors and other learners. The persona effect (Lester et al. 1997b; Lester et al. 1997c) was discussed as preliminary evidence for the positive impact of the presence of pedagogical agents on students' perceptions of their learning experience, but the mixed results of later research caution against its unconditional acceptance and suggest that the social behaviors of pedagogical agents need to be carefully designed to have educational value (Dirkin et al. 2005).

Current research emphasizes learning through and with computers but neglects the human-computer dimension (learning from computers). In contrast, Sims (1997a), from the perspective of a courseware developer, proposed a framework for analyzing learner-content interaction in computer-based learning environments (cf. Section 5.4.6). He identified a set of options for interactivity available to developers of educational software and integrated these interactive concepts into an engagement-control model of interactivity.

Sims' interactive concepts represent different levels of interactivity in the interaction of learners with computer systems and have been located along a continuum from low to high interactivity. While Schulmeister (2003; 2005) maintained that even at the highest level of interactivity, computer programs still do not have the human capacity for communication and social interaction, it was argued in Section 5.4 that the increasingly sophisticated verbal and non-verbal behaviors (cf. Chapter 10), natural language processing and dialogue capabilities (cf. Chapter 11), expertise (cf. Chapter 12), and emotional intelligence (cf. Chapter 13.4) of pedagogical agents enable them to participate in reciprocal and symmetric instructional exchanges with learners that involve meaningful actions and conversation.

6 Design

Design as an activity is the creative process of originating an idea and translating it into a blueprint or plan for an aesthetic and/or functional artifact. The term can also refer to the final blueprint or plan (a drawing, model, simulation, or other description), or to the result of its implementation (i.e. the appearance and functionality of the artifact produced). However, design is usually viewed as being distinct from implementation.

Design is a complex field which consists of various subdisciplines. Some examples are provided below:

- *Information design* defines, shapes and plans message content and presentation contexts to facilitate comprehension of the presented information.
- *Interaction design* is concerned with designing interactive experiences (cf. Chapter 5). It considers the following aspects: human affect and cognition (cf. Section 6.3 and Section 6.4), context of use (cf. Section 6.2), nature of the task, experience of the user, learnability of user interfaces, presentation of options, feedback on progress or errors, and recovery from failure in the interaction.
- *Interface design* creates the components of an interactive artifact through which it presents itself to the user and which allow the user to access its functionality (cf. Chapter 1.2).
- *Product design* develops artifacts that can be offered to a market. The process considers aspects of a product that include, among others, aesthetic appeal, usability, manufacturing, and market placement.
- *Visual design* creates visual experiences in any media to communicate a message, using design elements like point, line, form, texture, and color (McClurg-Genevese 2005b), and following principles like balance, contrast, proportion, rhythm, and unity (McClurg-Genevese 2005a).

Design is fundamental and ubiquitous. Design decisions affect nearly every part of our daily lives. The process of designing an artifact that fits seamlessly into and easily becomes at home in our world thus requires considerable research, thought, modeling, iterative adjustment and redesign. Not only are aspects like appeal, cost, image, safety, viability, etc. important, but good design always starts from the nature and needs of the people who will be using the artifact in a particular set of circumstances. In other words, the artifact should be designed with careful consideration for the *target group* of users (cf. Section 6.1) and the intended context of use (cf. Section 6.2). In general, the design and use of artifacts for people and by people in a particular context has a cognitive, an affective, and a social dimension. These three dimensions are discussed in Section 6.3, Section 6.4, and Section 6.5, respectively.

6.1 The User

A *user* is a person who makes use of a thing to achieve something, for himself or herself or for other individuals, alone or in cooperation or competition with other users, for work, learning, or pleasure. Things may be used for their inherent or natural purpose, or for a foreign purpose, which may not have been foreseen or intended.[51] In the field of human-computer interaction (HCI), a user is an individual who interacts with a computer through a user interface to carry out certain activities, such as writing an e-mail, playing a game, or debugging a program.

The term *affordance* refers to the perceived and actual properties of an artifact or system that indicate to users how to interface with it (Norman 2002a: 9). Good design creates intuitive affordances that meet existing user expectations. To achieve this, designers need to acquire knowledge about the people who will use the artifact or system. The target group of users may be more or less clearly defined in terms of the following (and other) attributes:

- *Gender* (masculine or feminine);[52]
- *Age* (e.g. children, teenagers, senior citizens, etc.);
- *Culture* (e.g. Asian vs. European vs. Latino-American);
- *Education* (e.g. academic degree, autodidact, (computer) literacy, on-the-job training, etc.);
- *Social class* (e.g. lower, middle, and upper class);
- *Occupation* (e.g. college students, office workers, retired or unemployed people, etc.);
- *Organization* (e.g. business, charity, government, military, university, etc.);
- *Intrinsic limitations* (cognitive, emotional, physical, and social);
- *Extrinsic limitations* (legal, monetary, social, spatial, and temporal);
- *Skills* (creative, cognitive, manual, and social);
- *Goals* (related to happiness, health, productivity, religion, safety, altruism, etc.);
- *Needs* (e.g. Maslow's Hierarchy of Human Needs, cf. Figure 30);
- *Expectations* (of interaction and performance, including consistency, effectiveness, efficiency, naturalness, predictability, and reliability);
- *Experiences* (disappointment, enjoyment, satisfaction, etc.);
- *Preferences* (e.g. input and output modes, interaction styles, language, brand, product design, etc.).

6.2 Context of Use

Context can be defined as the set of interrelated circumstances and conditions in which an object, event, process, statement, idea, linguistic unit, or other reality exists or occurs and which help to determine, specify, or clarify its meaning, function, or operation in that particular setting.[53] The term *context of use* refers to the situational factors that influence the use and usability (cf. Section 6.3) of a given artifact. Among other aspects, the context of use includes users (cf. Chapter 6.1) and other (artificial and human) agents, tasks, equipment (computer

---

[51] For example, people may use a lighter to light a candle (its inherent purpose) or to open a bottle (a foreign purpose). On the World Wide Web, guest books of web sites are often used as a kind of makeshift asynchronous chat facility, with users exchanging messages via guest book entries.

[52] Gender can be defined as the 'maleness' or 'femaleness' that individuals identify in themselves, project to others, or that other people perceive in them. It has to be distinguished from sex, which refers to the biological categories of male and female. While sex is determined before birth, gender is the result of a process after birth by which men and women are socialized to be masculine or feminine, learning and adopting from their social and cultural environment (Isbister 2006: 108).

[53] This definition is intentionally broad. In fact, the term 'context' has a wide range of meanings in different domains and disciplines. For example, it means different things in artificial intelligence, contemporary art, literature, computer operating systems, and sitcoms.

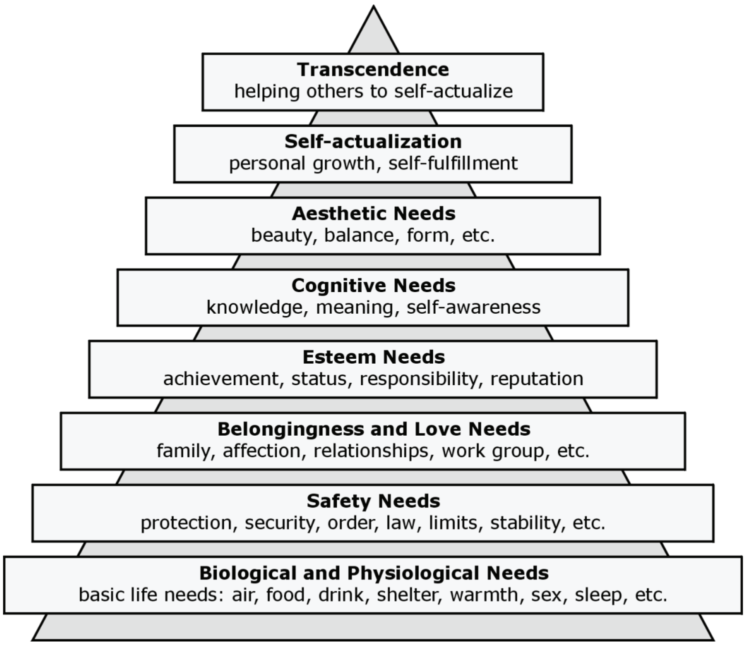

Figure 30. The Hierarchy of Human Needs was originally developed by the psychologist Abraham Maslow (1908–1970) in the 1940s and later extended by him and others (Maslow 1943; Maslow 1970). The diagram shows the eight-stage model from the 1990s (Chapman 2002). The hierarchy is a psychological model of human motivation, which states that humans meet their basic needs (physical and emotional well-being) first before they seek to satisfy higher order needs (knowledge, beauty, personal development, helping others).

hardware, software, networks, materials, machines, and tools), and the physical, sociocultural, and organizational environments in which the artifact is used. Physical context factors include space, lighting, temperature, noise level, and so on. Examples of sociocultural factors include the customs, fashions, norms, roles and role allocation (cf. Chapter 7.14), and other aspects of the target culture (cf. Chapter 7.15). The social network, management and organizational pressures, and processes of learning, teaching, and working are examples of organizational factors. Furthermore, time and its passage are another important aspect of the context of use. Not only does the temporal context affect the use and usability of an artifact (e.g. the artifact cannot be used before or after a certain point in time), but the context of use may also change with the passage of time. Artifacts that process natural language (cf. Chapter 11) additionally have to consider the *linguistic context*, i.e. the set of factors (provided by the surrounding text or utterances) which systematically determine what form a linguistic unit (a speech sound, word, phrase, sentence, turn, paragraph, text, or dialogue) takes, what it means, how appropriate it is, or how it should be translated (Arnold et al. 1994: 201). Different aspects of the linguistic context are discussed in Chapter 11 (see also Franklin (2003: 194–196) for a brief discussion of selected examples). A distinction is often drawn between the linguistic context and the *non-linguistic context*. The latter includes knowledge and beliefs (cf. Chapter 12.1) shared by the speaker/writer and hearer/reader that are not internal to language but relate to external domains of experience.

Context awareness is the capability of artifacts to sense and act on relevant situational factors and their changes over time in a timely fashion, allowing them to provide pertinent information and/or services in a given context and to adapt to each new or modified context of use.[54] By taking a series of snapshots of the context across a time span, it is possible to create a *context history* (Chen & Kotz 2000), which can be used as a memory of contexts in which previous operations or interactions with the user took place. This memory can help artifacts to predict and thus better adapt to future contexts and interactions.

Careful consideration of the context(s) in which an artifact will be used by the target group of users from the early stages of design to the final field testing stages makes it possible to develop products which are better suited to the needs of the situation and the users, resulting in better usability and performance and hence a higher degree of user satisfaction. Therefore, usability testing, prototyping sessions, focus groups, end-user studies, and other types of user involvement in the design process (cf. Chapter 7.13) have to be carried out in the artifact's intended context of use ([INT 46], cf. Chapter 15).

For software agents, the context of use comprises the situational factors of the virtual and physical environment in which an agent is embedded (cf. Chapter 3). The virtual environment, consisting of computer hardware, software systems, and networks, is the agent's habitat while the user is situated in the physical world (although the distinction between the virtual and physical realms is becoming increasingly fuzzy, with more and more people spending larger and larger portions of their lives in *massively multiplayer online role-playing games (MMORPG)*, virtual communities, and other *synthetic worlds*[55] (Castronova 2006)). The agent's ability to perceive and act covers the virtual environment and may even extend to the physical environment if the agent is equipped with the appropriate sensory and actuatory hardware (e.g. cameras, microphones, mechanic bodies, etc.). Contextual information is collected via the agent's sensors (cf. Chapter 3 and Chapter 4.2). Time and resources are usually limited; hence, the agent has to separate relevant from irrelevant contextual information when processing the context, i.e. extract the situational factors pertaining to its goals and tasks. Furthermore, the complexity and dynamics of environments can easily exceed the capabilities of current sensory devices and algorithms, in particular those for perceiving the physical world, so agents have to allow for a certain *error rate* in their perceived input (and subsequent processing of that input).

Context processing is important for various aspects of the agent's operation, including disambiguating inputs, interpreting relationships, and generating outputs. To be able to use the contextual information extracted from the sensor input for its internal decision-making and planning processes, the agent has to create and maintain a representation of the (changing) context over time. This may involve combining lower-level contexts with other contexts or information into higher-level contexts (Chen & Kotz 2000). The result is a *context hierarchy*, which breaks down the context into successively more specific subcontexts (Schmidt 2002).

6.3 Cognitive Aspects of Design

The cognitive dimension of design centers on the concept of *usability* (Norman 2002a), which describes the quality of users' experience when they interact with a human-made artifact to achieve a particular goal. The International Organization for Standardization, widely known as

---

[54] More precisely, this is what Chen and Kotz (2000) call *active context awareness*. In addition, they refer to *passive context awareness* as the ability of systems to dynamically present new or updated contexts to users and/or to store contexts for later retrieval.

[55] According to Castronova, a synthetic world is "an expansive, world-like, large-group environment made by humans, for humans, and which is maintained, recorded, and rendered by a computer" (Castronova 2006: 11).

ISO, defined usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO 1998). Users may seek to achieve their own goals or the goals of others, for work, learning, or pleasure. Goal achievement is characterized by the quality of the result, the time and effort that had to be expended in the process of achieving the goals, and the user's sense of accomplishment. Finally, a given product is always used under a set of specific conditions in a real-life or laboratory setting (cf. Section 6.2).

In human-computer interaction, usability is a measure of how effectively, efficiently, and satisfactorily a person can navigate a user interface, find information on it, and achieve his or her goals. A combination of factors affects this experience ([INT 47]):

- *Ease of learning*. How fast can a novice user learn how to use the system?
- *Efficiency of use*. How much time and effort do experienced users require to carry out tasks?
- *Memorability*. How easily can the user remember how the user interface works?
- *Error frequency and severity*. How frequent and serious are user errors in the interaction, and what error recovery mechanisms are available?
- *Subjective satisfaction*. How enjoyable is it for people to use the system?

Norman (2002a: 13–29 + 52f.) suggested four principles for usable and understandable designs. First, designers should *make things visible*, i.e. always show clearly what the current state of the system is and what the user can do. Second, they should provide *a good conceptual model* for the user (pp. 189f.). The design and implementation of a system is guided by the *design model*, i.e. the conceptualization (or conceptual model) of the system in the minds of its designers. As users interact with the system, they build their own mental model of beliefs about the system and its operation (the *user's model*) (Norman 2004: 75). This model helps them to understand the system, which includes explaining and anticipating its behavior. Designers expect the user's model to be congruent with their design model. However, users can only learn about the design model indirectly through the *system image*, which is created by the way the system appears, operates, and responds to user actions, as well as any written material (advertisements, documentation, test reports, etc.) describing it (Norman 2002a: 190). Hence, designers must portray their system in a way that is consistent with the underlying design model and makes its operation transparent (Norman 2004: 76), thus allowing users to build the appropriate mental model. Figure 31 visualizes the relationships between the design model, the user's model, and the system image.

Norman's third principle requires designers to provide *good mappings*, i.e. relationships between controls, affected entities, and results ([INT 48]) that make it easy for the user to map from actions to results, from controls to effects, and from the internal system state to observable system behavior (Norman 2002a: 53). Mappings are preferably *natural*, which means that they exploit physical analogies (e.g. controls are moved in the same direction as intended for the object they affect and cultural standards (e.g. screws are tightened clockwise) (p. 23).

Norman's final principle requires systems to give *feedback*, i.e. to inform the user about the actions performed by the system and their results (Norman 2002a: 27). Feedback should be timely (provided when the user needs or expects it), relevant (give the user the information they want or need), clear (be free from ambiguity), sufficient (not give more information than necessary), and (possibly) personalized (adapted to the user's characteristics, cf. Section 6.1).

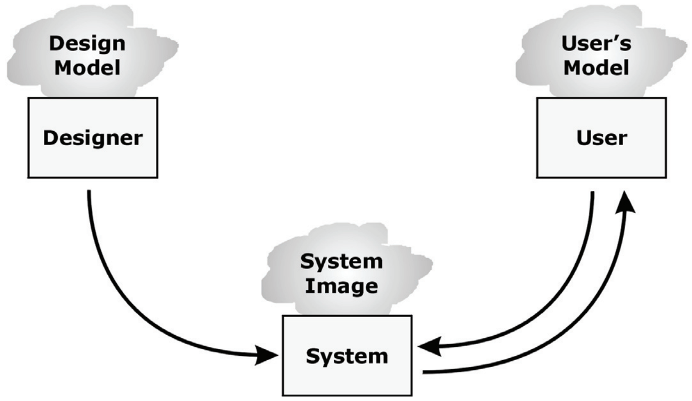

Figure 31. Three aspects of mental models: the design model, the user's model, and the system image. Adapted from Norman (2002a: 190, Figure 7.1).

6.4 Affective Aspects of Design

Usability design is rooted in the cognitive sciences.[56] *Cognition* is an information processing mechanism that gives individuals the ability to interpret, reflect on, and remember things in their environment (Norman et al. 2003: 38). *Cognitive artifacts* are physical objects or software applications which are designed to support cognitive processing of humans or to enhance their cognitive abilities, by facilitating mental processes of remembering, thinking, and problem solving. Examples include calendars, to-do lists, diagrams, word processors, and pedagogical tutor agents.

Cognition is tightly coupled with another information processing system: *affect*,[57] which continuously evaluates and judges ('appraises,' cf. Chapter 13.1.2.1) the environment as it is perceived by and concerns the individual, providing immediate positive (e.g. 'desirable,' 'good,' or 'safe') or negative (e.g. 'undesirable,' 'bad,' or 'dangerous') assessments of things in the world (Norman et al. 2003: 38). Many believe cognition and affect to be opposites, regarding the former as cool, human, and logical while describing the latter as hot, animalistic, and irrational (Norman 2004: 7). In fact, both systems are functionally inseparable and mutually dependent (cf. Chapter 13.2.1). Affect influences cognition, and cognition drives a number of affective states. Importantly, affect directs and focuses attention on entities and states of affairs that individuals have appraised as salient to achieving their goals or satisfying their needs (Nass & Brave 2005: 77). Furthermore, affect influences cognitive style (e.g. in problem solving) and task performance (easy tasks become more difficult, difficult tasks become easier), which has important implications for design (Norman 2002b), as shown in Table 7.

---

56 The cognitive sciences (Wilson & Keil 2001; Harnish 2002; Nadel 2003) combine the analytical fields of psychology, computer science, human factors, and engineering. They study topics bearing on the nature and operation of the human mind and intelligence.

57 In emotion research, *affect* is a head term for various conditions including, among others, emotions, moods, and preferences (Ortony et al. 2005: 174). Affect varies with respect to *valence* (the pleasant/unpleasant dimension: positive or negative) and *arousal* (the degree of intensity: very high/high/medium/low/very low) (Lisetti 2002).

Table 7. The impact of affect on cognition. Compiled from Norman (2002b; 2004).

Affect	Effects	Implications for Design
Negative	Focuses thinking. '+': Improved concentration; '–': Tunnel vision.[58] Simple tasks become more difficult.	Highlight function and minimize distractions, bottlenecks, and irritations in stressful contexts.
Positive	Broadens thinking. '+': Enhanced creativity; '–': More easily distracted. Difficult tasks become easier.	Emphasize aesthetics in neutral or positive situations; attractive designs increase positive affect, creativity, and tolerance for minor difficulties.

Finally, affect has an effect on people's *judgment*. The number of options in decision making is reduced by rapid assessments from the affective system. But affect also functions as a filter through which individuals assess any entity, situation, or event, even if it has no relation to their current affective state (Nass & Brave 2005: 78). People in a good mood are more likely to give positive assessments and follow recommendations. However, while they are more willing to take risks when making *hypothetical* decisions, they are more cautious in actual risk situations (Isen 2000).

Anxious people are more focused and thus require relevant information to be accessible and easily visible at all times, as well as clear and unambiguous feedback from an artifact about its operations. In contrast, people in a good mood are more creative and tend to be more willing to tolerate minor shortcomings of an artifact, in particular if they perceive it as fun and enjoyable (Norman 2004: 26). Isen (2002) found substantial evidence that experienced mild positive affect generally has positive effects on negotiation processes and their results; generosity and social responsibility (helping others and volunteering); self-efficacy; self-motivation to accomplish tasks; and openness to and flexible processing of new information (Burleson 2005: 448).

Technologies with the ability to induce positive affect in users, in particular when they sense (or reason) that an activity or experience negatively impacts users (cf. Chapter 13.4.1), may improve users' emotional well-being, intrinsic motivation (cf. Chapter 13.2.2), performance, and creativity. For example, technologies could give users the opportunity to volunteer and encourage them to apply their skills in meaningful ways in the context of tasks that are challenging but also interesting and/or important (preferably both).

The fact that affect and cognition are inseparable and indispensable for behavior indicates that the design of artifacts should not only be influenced by considerations of usability but also involve a strong, at least equally important, *emotional* component, an idea discussed in detail by Norman (2004). Hence, artifacts should be both usable (cf. Section 6.3) and pleasurable. "Attractive things work better" (Norman 2002b) because they make people feel good, which promotes curiosity and creativity and thus the ability to find alternative solutions to problems (Norman 2004: 26). What is true of artifacts has also been demonstrated to hold for people. There is plenty of research supporting the so-called *halo effect*: good-looking people with a healthy, symmetrical face and body, clear skin, straight profile, and other (culture-specific)

[58] *Tunnel vision* is a state of mind where people have a single-minded focus, with the result that they become blind to other options that would be obvious to them under normal circumstances (Norman 2004: 28).

Table 8. Levels and design features in the model of cognitive and affective design (Norman 2004).

Processing Level	Design Level	Salient Features of Artifacts
Reflection	Reflective	Self-image, personal satisfaction, memories
Routine	Behavioral	Functions, performance, usability
Reaction	Visceral	Appearance, texture, sound, taste, smell

attractive aspects (Isbister 2006: 7ff.) are ascribed more positive traits than less attractive people, including biological qualities like good genes, developmental and hormonal health, and a strong immune system (Green et al. 2008: 2457) as well as individual and social qualities such as independence, sociability, intellectual capability, trustworthiness, and appeal (Jones et al. 1978; Brigham 1980; Langlois et al. 2000; Green et al. 2008: 2459). Physical attractiveness contributes significantly to the social influence of both humans and artifacts: someone or something that is visually attractive to a certain target audience is likely to be more persuasive than people or artifacts that are perceived as unattractive (Fogg 2003: 92ff.). The fact that people tend to subconsciously assume all these things about others based on their looks can be exploited in the design of the appearance (cf. Chapter 9.2) of embodied agents (Gard 2000; Isbister 2006: 7ff.). However, designing appealing facades is not the lesson here: in good designs, appeal and usability (cf. Section 6.3) are in balance (cf. Chapter 7.3).

Norman presented a model of design whose levels correspond to a three-level model of human behavior where processing at each of the three levels assigns meaning (cognition) and value (affect) to entities, situations, and events in the world (Norman 2004: 25; Norman et al. 2003: 39). Table 8 gives an overview of the model. The relationship between the processing levels is shown in Figure 32. The following discussion of the three-level model of affect and cognition and the corresponding design levels draws primarily on the accounts by Norman et al. (2003) and Norman (2004).

The *reaction level* contains automatic, pre-wired behaviors that allow an organism to quickly assess whether something is good or bad, safe or dangerous, etc. and to respond to perceived dangers or problems at once, by interrupting processing at higher levels, increasing arousal, and signaling the motor system to initiate or prepare an immediate response. Reactive processes are determined by genetic mechanisms and can be enhanced or inhibited by control information from higher levels. Environmental stimuli automatically trigger responses, with no reasoning involved. Table 9 lists various cues from the environment that can produce automatic positive or negative affect.

The *visceral level* of design considers users' automatic judgmental reactions to these cues and accordingly designs artifacts sending affective signals that elicit a favorable initial response from the members of the target group of users. The major elements of visceral design include appearance, texture, sound, taste, and smell. At the visceral level, the goal is to create artifacts that users perceive as, for example, 'pretty,' 'cute,' 'fun,' or 'tasty,' in general as 'attractive.' Humans have certain genetically determined preferences with regard to shape, symmetry, taste, smell, color, sound, touch, setting, and appearance and expressions of the human body, which are listed in the left column of Table 9. These features tend to produce positive affective reactions across cultures; however, designers also need to consider the influence of culture and the passage of time on what is considered as attractive and what is not (cf. Chapter 7.15.3).

The *routine level* is the locus of default expectations as well as well-learned, routine behaviors. This unconscious mid-level controls everyday human behavior, in particular most motor skills (including language production). At this level, the individual is aware of its environment but not of his or her own individuality (Ortony et al. 2005: 179). Processing at the

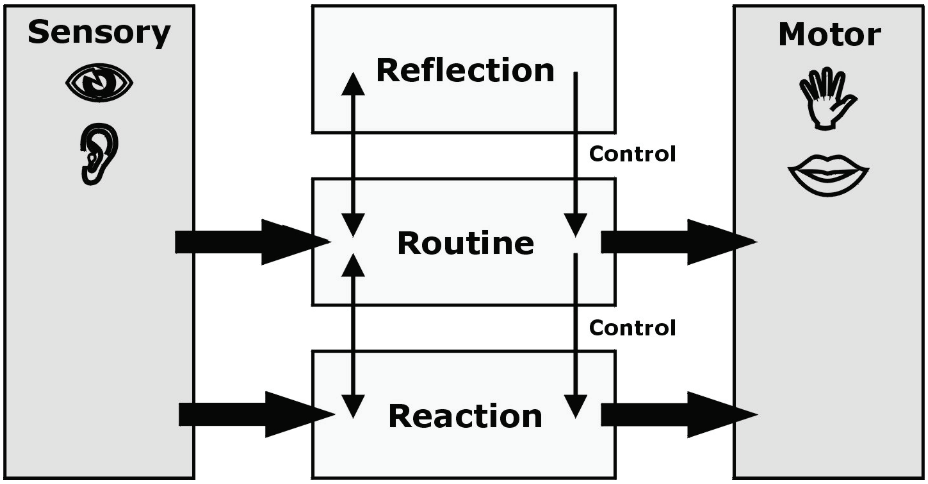

Figure 32. A model of affect and cognition. Adapted from Norman (2004: 22, Figure 1.1).

routine level is influenced by inputs from the sensory system, information from the reaction level, and control signals from the reflection level. The routine level exercises control of the reaction level by enhancing or inhibiting the responses of the lower level. Furthermore, it notifies the reflection layer when it detects something that deviates from its default expectations.

Behavioral design targets the routine level of processing. It is concerned with the pleasure and effectiveness associated with the use of an artifact. Therefore, it emphasizes learnability as well as efficiency and effectiveness of use in design. At this level of design, artifacts have to possess features that endow them with function, understandability (or comprehensibility, cf. Chapter 7.7), usability (cf. Section 6.3), and physical feel.

Function has to do with what an artifact does, what it is 'good' for, what purpose or role it has in a specific context of use (cf. Section 6.2). If the artifact has no discernible function or does not provide the function users need in a given context, it will be of little interest to users. As a result, designers have to learn about users' needs through early and continuous involvement in the design process (cf. Chapter 7.13). What exactly people's needs are (cf. Figure 30) is often not easy to determine, though. Users may not even be aware that they have a particular need which the artifact could serve.

Understandability is a design feature that enables users to form the correct mental model of what an artifact does and how it works from the system image portrayed by the artifact. From the designers' point of view, the user's mental model is 'correct' if it is congruent with their own conceptualization of the artifact (cf. Section 6.3). To make sure that the image projected by the artifact allows users to build the desired mental model, user testing with successively more refined prototypes is necessary. The quality of a system image is indicated by the degree to which it makes transparent the purpose and operation of the artifact as well as its ability to give explanations of decisions, actions, and results occurring while someone is using the artifact, if necessary, in order to make them comprehensible to the user (cf. Chapter 7.7).

Artifacts may serve the required function and may be comprehensible to users, but they also have to be designed in such a way that they allow for effectiveness, efficiency, and satisfaction of use. The concept of *usability* was discussed in detail in Section 6.3. Here, it should be added that interaction with everyday artifacts should not require extensive prior training or involve complex sequences of actions. Instead, the interaction should be as natural as possible, taking

Table 9. Factors creating positive or negative affect. Adapted from lists in Norman (2004: 29f.).

Automatic Positive Affect Produced By	Automatic Negative Affect Produced By
Environmental Conditions	
• Temperate climate	• Extreme hot or cold • Darkness
Animate and Inanimate Things	
• Symmetrical objects • Sounded, smooth objects	• Looming objects (seemingly about to hit the observer) • Sharp objects • Snakes and spiders
Sense Impressions	
• Sweet tastes and smells • Bright, highly saturated hues • Soothing sounds • Simple melodies and rhythms • Harmonious music and sounds • Caresses • Rhythmic beats • Sensuous feelings, sounds, shapes	• Sudden, unexpected loud sounds or bright lights • Extremely bright lights or loud sounds • Harsh, abrupt sounds • Grating and discordant sounds • Rotting smells, decaying foods • Bitter tastes
Places	
• Warm, comfortably lit places	• Heights • Crowds of people • Empty, flat terrains (deserts) • Crowded dense terrains (jungles or forests)
The Human Body	
• Smiling faces • Attractive people	• Deformed human bodies • Human feces (and their smell) • Other people's body fluids • Vomit

advantage of what users know without training, by virtue of being human, such as their ability to engage with others in face-to-face conversations (cf. Chapter 1.2).

Finally, for physical artifacts, *tangibility* (weight, texture, and surface) is an important design consideration. The physical feel and touch influence people's liking or disliking of an artifact and are critical to their assessment of its behavior.

The *reflection level* is the level of self-awareness, at which the mind contemplates its own representations and operations. This level involves high-level cognitive functions, including conscious thought, full-fledged emotions (cf. Chapter 13.1), metacognition,[59] self-reflection,

[59] Burleson, referencing the definition by Blakey and Spence (1990), defined *metacognition* as "the process of thinking about thinking, understanding what is known and not known. It involves: connecting new information to former knowledge; deliberately selecting and monitoring thinking strategies; and evaluating these processes" (Burleson 2005: 443). Self-regulating one's own cognitive

and learning (Ortony et al. 2005: 179). The reflection level is detached from sensor input and behavior control. However, it monitors, thinks about, and attempts to bias the operation of the routine level. On the other hand, its own processing is influenced by inputs from the lower levels.

Reflective design relates to the reflection level of processing. It is concerned with the (inter-) personal and sociocultural meaning and message of (using) a particular artifact, i.e. with how its users rationalize and intellectualize what it means to themselves (or to others) and what it communicates (about themselves) to other people in a given cultural context. At this level, designs should seek to elicit positive answers to questions such as the following:

- Does owning or using the artifact enhance the user's self-image[60]?
- Does the artifact send a message to others which the user approves of?
- Can the user make memories (and stories to tell others) with and through the artifact?
- Can the user form a lasting relationship with the artifact (cf. Chapter 7.16)?

The visceral, behavioral, and reflective levels are interwoven and permeate any design (Norman 2004: 6). However, particular designs may put different emphasis on the individual levels. Some artifacts may have visual appeal but lack usefulness or usability; some may be intended for one-time use only, whereas others are designed with the potential to become life-time companions to their users. Furthermore, different users or groups of users may have varying preferences with respect to the three levels in different artifacts or even the same artifact. Some users value appeal while others favor utility and still others appreciate meaning and message. Overall, designers need to be sure about what level(s) of processing (reaction, routine, and/or reflection) they want to appeal to in order to achieve their goals with a given artifact (sales, brand recognition, user satisfaction, learning gains, relationship building, etc.), and they need to learn as much as possible about the future users of the artifact they are designing (Norman 2004: 39), preferably by involving them early in and throughout the design process.

6.5 Social Aspects of Design

In addition to usability and attractiveness, design has to consider the social responses that artifacts elicit in users. The *Media Equation* (Reeves & Nass 1998) is a theory of the interaction between humans and media (computers, television, and new media) which is based on the idea that people's interactions with media are "fundamentally social and natural" (Reeves & Nass 1998: 5). That is, people interact with media as if they were real people and places, applying the same social and physical rules about face-to-face interaction with other people and the way the physical world works: "media equal real life" (p. 5). Reeves, Nass, and their colleagues found that *all* people are subject to the Media Equation, regardless of age, culture, education, and experience with technology,[61] and that their social responses to media are *automatic* and

processes requires an awareness of those processes ([INT 49]) and the ability to consciously reflect on and manipulate mental knowledge representations.

[60] A person's *self-image* is the mental conceptualization which that individual has created of himself or herself, including views of his or her appearance, abilities, and personality, particularly in relation to others and their perceptions and opinions of the individual.

[61] There are other empirical studies on social attributions to computing technology that are more careful with claims of universal susceptibility. For example, Johnson et al. (2006), based on a model of the

unconscious (Reeves & Nass 1998: 252), i.e. “deeply instinctual and not available to conscious introspection” (Mishra et al. 2000: 3), in short ‘mindless’ (Nass & Moon 2000). While counterintuitive,[62] the Media Equation is supported by substantial empirical research (Nass & Steuer 1993; Nass & Sundar 1994; Nass et al. 1994; Nass et al. 1995; Moon & Nass 1996; Nass et al. 1996; Fogg & Nass 1997; Nass et al. 1997a; Nass et al. 1997b; Reeves & Nass 1998; Takeuchi et al. 1998; Nass et al. 1999; Nass & Moon 2000; Nass et al. 2000a; Takeuchi et al. 2000; Fogg 2003; Ferdig & Mishra 2004; Lee & Nass 2004; Nass & Brave 2005), which has produced surprising results. Some of the empirical findings reported include:

- People are polite to computers, whether the computers are using text or voice (Nass et al. 1999).
- Humans automatically attribute gender to a voice, determine whether the assigned gender matches their own, and show social identification if it does, conforming more to voices with the same gender and regarding them as more trustworthy (Nass & Brave 2005: 16f.).
- People treat computers speaking with female voices differently from those with male voices. Both men and women are more influenced by positive evaluation from a computer with a male voice than from a computer with a female voice, feel that dominance by a female-voiced computer is unbecoming, and believe that a computer with a female voice is more knowledgeable about stereotypically feminine topics, whereas a male-voiced computer knows more about stereotypically masculine topics (Nass et al. 1997a).
- People respond to different voices as different social actors. The same voice on different machines is treated as the same social actor (Nass & Steuer 1993).
- People perceive different computer-generated voices as different individuals. Like human voices, synthetic voices are indicators of individuality (Nass & Brave 2005: 103). Consistent with the multiple-source effect,[63] several synthetic voices are more persuasive than a single synthetic voice, and they lead to the experience of greater social presence (Lee & Nass 2004).
- People recall more information from a foreign-language tutorial involving a computer reading out instructions with a *native* accent than from a tutorial featuring a computer in the same role but with a *non-native* accent because they apply the stereotypical assumption that “a native speaker knows better” and thus pay more attention to the tutorial (Alvarez-Torres et al. 2001).
- People are influenced by both sincere praise and flattery (insincere praise) delivered by computers and respond to them in similar ways as they do when other people praise or flatter

---

computer as a social actor described by Marakas et al. (2000), found that the tendency of people toward social attribution is influenced by their core self-evaluation (i.e. how they subconsciously see themselves with respect to their view of the environment, confidence and sense of control), what they believe about the social role of computers, and the characteristics of the user interface. However, their results still indicated that “a significant portion of our society do indeed believe that computers have social agency and respond to computing technology accordingly. In fact, this propensity is actually magnified when exposed to a distinctly social interface” (Johnson et al. 2006: 457).

[62] Most people would reject the idea that they respond socially to media. In fact, the participants in the studies conducted by Nass, Reeves et al. uniformly were in denial of the specific social behaviors which they actually had shown during the experiments (e.g. Nass & Moon 2000).

[63] According to the *multiple-source effect* (also known as the *source-magnification effect*), multiple people giving the same opinion or statement are more convincing than a single source (Harkins & Petty 1981; Harkins & Petty 1983; Harkins & Petty 1987; Nass & Brave 2005: 100).

them. In particular, they feel better about themselves, their performance, the human-computer interaction, and computer-generated evaluations when they receive sincere or insincere praise from a computer (Fogg & Nass 1997).
- People like a praising computer more than one that criticizes.[64]
- People think that self-praising media are less likeable and less competent than media that receive praise from others. Critical media are liked less when they criticize others than when they criticize themselves, but they are also perceived as more competent.
- People respond to praise or blame provided by a computer but seem to take the feedback at "face value," unwilling to reason more deeply about the intentions and reasons behind the computer's responses (Mishra 2006).
- People imbue computers with a personality (Nass et al. 1995), and their social responses to computer personalities are similar to their responses to human personalities. Furthermore, people are more attracted to a computer personality that is similar to their own (in terms of dominance vs. submissiveness) compared to one that they perceive as different (Moon & Nass 1996).
- Images of faces that appear close make viewers react more strongly than distant images.
- People respond differently to 'generalist' and 'specialist' TV news programs.
- People can be convinced to accept computers as teammates, just by telling them that they are interdependent with the computer. The effects of perceived team affiliation are the same for human-computer teams as they are for all-human teams: human team members think that they are more similar to the computer and more cooperative. They are also more likely to be influenced by the computer; think that information provided by the computer is friendlier and of higher quality; and are more willing to conform to the computer (Nass et al. 1996).
- People regard a computer itself as the source of information, not the programmer.
- People show the same types of social reciprocal behaviors to computers as they exhibit in their interactions with other humans. Reciprocal behaviors are modulated by the social norms of the individual's cultural background (Takeuchi et al. 1998; Takeuchi et al. 2000).
- People can feel betrayed by computers and may retaliate with spiteful behavior, like they do toward humans. They tend to interpret the behavior of computers as intentional, accusing them of deception, lying, and trickery (Ferdig & Mishra 2004).

People behave socially toward media because they know from their experience in the real world that sophisticated social behavior involving the use of language is confined to humans and that whatever object they perceive is real, and they apply this knowledge to mediated experiences (Reeves & Nass 1998: 12). Constantly looking for patterns in and trying to make sense of what (or who) is around them, whether mediated or real, it comes naturally to humans to categorize and stereotype media in the same way as other humans (Mishra & Wojcikiewicz 2002). Entities that *seem* to be real people or locations are automatically *perceived* as real, including simulations of social actors and real-world objects, and these perceptions influence behavior (Reeves & Nass 1998: 12). It appears that this *intentional stance* (Dennett 1987), "the strategy of interpreting the behavior of an entity (person, animal, artifact, or the like) by treating it as if it were a rational agent that governed its 'choice' of 'action' by a 'consideration' of its 'beliefs' and 'desires'" (Dennett 2001: 412), is a matter of instinct rather than conscious control and is not influenced by the individual's age, experience, or expertise (Weizenbaum 1976; Turkle 1984; Reeves & Nass 1998; Mishra & Wojcikiewicz 2002).

[64] Findings without bibliographical references were reported by Reeves and Nass (1998).

Overall, the Media Equation has a number of important implications for the design of media (Reeves & Nass 1998: 252ff.):

- Only *minimal cues* are required to elicit social responses. For example, on-screen text and simple line drawings can create the perception of a personality.
- Users tend to be influenced by their perceptions of reality rather than reality itself. Hence, media do not have to *be* identical (with respect to their qualities) to the real-world entities they portray; it suffices that they are *perceived* that way (cf. Chapter 7.2). For example, perceived intelligence or personality may matter more than actual intelligence or personality in terms of human responses.
- People respond to what they see, hear, and interact with, i.e. to what is *immediately present*. For example, they assess intelligence based on appearance and language ability. They also regard an animated on-screen character delivering a message as the source of that message.
- Users prefer *simplicity* and *predictability*: they feel more comfortable when there are fewer choices to make[65] and when they know what they can expect. Media can thus help users by offering them freedom *from* (rather than *of*) choice and acting predictably and reliably in interactions.
- Media should adhere to the rules governing social relationships and (inter-) actions in the real world that are already familiar to users.

The Media Equation provides the basis for a line of research in the field of human-computer interaction which contends that people tend to interpret their interactions with computers in the same way as their interactions with other humans, provided that the computer as their partner in the interaction exhibits sufficient *social cues* (Louwerse et al. 2005: 694). Table 10 lists five primary types of social cues which users perceive as indicators of the presence of a living entity. In the context of multimedia learning (cf. Chapter 2.2.2), Mayer et al. proposed the *social agency theory*, according to which

> *social cues in a multimedia message can prime the social conversation schema in learners. Once the social conversation schema is activated, learners are more likely to act as if they are in a conversation with another person. Thus, at least to some extent, the social rules of human-to-human communication come into play. (Mayer et al. 2003: 419).*

Mayer et al. found in two experiments that the presence of social cues (specifically voice) activated a social conversation schema in learners, leading them to perceive the computer as a conversation partner rather than a presenter of a multimedia message. As a result, learners made more of an effort to understand the message, which, in turn, resulted in deeper learning (Mayer et al. 2003: 420 + 424).

If computers exhibit appropriate social cues, learners are more likely to accept them as social partners in the learning process. The *social-cue hypothesis* predicts that the presence of social cues, such as facial expressions or human voices, in a computer-based lesson will improve students' learning and cause them to rate the lesson more favorably (Moreno et al. 2000). One way to integrate social cues into computer-based lessons is to have them portrayed by animated pedagogical agents. Several studies found positive effects for the voice of an agent (Moreno et

---

[65] For example, concerning the question whether users should be given a choice from among a set of voices in an interface, Nass and Brave argued that while this may invite a close relationship between the user and the voice of his or her choice, in particular if the user can interact with the voice (for instance if it belongs to an agent), the number of available voices should nonetheless be limited in order to avoid overwhelming the user with too many options (Nass & Brave 2005: 110f.).

Table 10. Primary types of social cues. Adapted from Fogg (2003: 91, Table 5.1).

Type of Social Cues	Examples
Physical	Face, eyes, body, posture, movement, voice
Psychological	Preferences, humor, personality, emotions, empathy, ruefulness
Language	(Spoken) language understanding and generation, conversational behavior
Social dynamics	Turn-taking, cooperation, politeness, reciprocity
Social roles	Doctor, teammate, opponent, teacher, pet, guide

al. 2000; Moreno et al. 2001; Mayer et al. 2003; Louwerse et al. 2005) and presenting the agent's words in a personalized conversational style (Moreno et al. 2000), whereas the visual presence (image) of an animated pedagogical agent had no effect on learning and ratings (Moreno et al. 2000; Moreno et al. 2001). However, social agency is not necessarily promoted by adding more social cues to an interface. In fact, Louwerse et al. found that any social cue can induce users to view the computer as a social partner (Louwerse et al. 2005: 702). If the voice is sufficient by itself, the image of an agent has little effect (Moreno et al. 2001; Craig et al. 2002).[66]

6.6 Summary

Design is the complex creative process of coming up with an idea and translating it into a plan for an artifact that meets aesthetic and/or functional requirements. Since artifacts have to be designed with careful consideration for their target group of users and their intended context of use, this chapter began with a discussion of these two aspects in Section 6.1 and Section 6.2, respectively.

The remainder of the chapter was concerned with the cognitive, affective, and social dimensions of the design of artifacts. The discussion of cognitive design aspects (cf. Section 6.3) focused on the concept of usability, which describes the extent to which an artifact can be used effectively, efficiently, and satisfactorily by a particular group of users in a specific context of use. Norman's four principles for creating usable and understandable designs were discussed (make things visible, provide a good conceptual model, provide good mappings, and give feedback). Mental models of an artifact or system have three aspects, which need to be harmonized: the design model (how designers conceptualize the system), the user's model (how the user understands the system and its operation), and the system image (how the system presents itself).

The affective dimension of design was the topic of Section 6.4. The discussion emphasized that cognition and affect, while seemingly opposite, are closely interconnected and influence each other, and went on to argue that the emotional side of the design of artifacts deserves

[66] It should be noted that in none of the experiments cited there was true conversational interaction between the pedagogical agents and learners. The agents merely presented information through voice, gesture, etc., but learners could not reciprocate using speech, gesture, or facial expressions. Hence, it seems that these presenter agents (cf. Chapter 12.3.12) did not tap the full potential of embodied pedagogical agents because they did not engage the learners in (multimodal) instructional conversations.

careful consideration. In other words, artifacts should be both usable and pleasurable for users because “[a]ttractive things work better” (Norman 2002b).

The third dimension of design is concerned with the social responses that artifacts elicit in users. The discussion in Section 6.5 focused on the Media Equation (“media equal real life”), which has become a particularly influential theoretical framework for explaining the ubiquitous, ‘mindless’ social responses of people to computers, television, and other media. Support for the Media Equation comes from a large body of empirical studies, which have produced surprising, even counterintuitive results, some of which were presented in Section 6.5.

7 Design Principles

Design is a means-to-ends process, i.e. it has a definite goal in view: an artifact that represents a solution to a given problem. The essence of design is thus problem solving (Manica 2005). Design problems are real-life problems with many solutions, some of which are more usable (Norman 2002a), enjoyable (Norman 2004), and socially adept (Reeves & Nass 1998) than others (cf. Chapter 6).

Gut instinct and inspiration play an important part in the process of solving design problems; however, these things alone cannot prevent mistakes that introduce flaws in the design, which in the worst case can lead to an unbalanced or incomplete product. A better strategy bases the design on principles which ensure that the outcome meets the user's requirements and is in line with knowledge about effective design. The following sections present a set of principles for designing embodied interactive pedagogical software agents.[67] While the focus is on agents that facilitate learning, the principles discussed can also be applied to the design of embodied interactive agents for other areas (e.g. computer games, e-commerce, and human-computer interfaces, cf. Chapter 3.2).

7.1 Added Value

One objection that has been made against animated agents for education (and other domains) is that adding another (eye-catching) object to a multimedia learning environment introduces a source of distraction for the learner, particularly if the agent exhibits behaviors which are meant to enhance its own believability (e.g. toe-tapping, head-scratching, and cart wheeling, cf. Chapter 10.2) but serve no purpose with respect to what the learner is doing at the moment (cf. Chapter 3.2.5). While recent empirical studies have shown that the presence of an animated agent is not detrimental to student learning (Craig et al. 2002), but apparently quite the opposite (Lester et al. 1997b; Lester et al. 1997c; Moreno et al. 2000; Atkinson 2002; Babu et al. 2007; Kim et al. 2007), it is clearly not enough for the agent to be cute and entertaining: rather, it must *add real value* to the learning experience. A number of studies (Moreno et al. 2000; Atkinson 2002; Babu et al. 2007) indicated that this added value was provided by the agent's ability to *communicate* verbally and non-verbally with the learner. Therefore, it seems that pedagogical agents should be designed to facilitate instructional conversations between agent and learner by including advanced speech and language processing technologies (Jurafsky & Martin 2008) and by giving the agent the ability to use its face, hands, and body in human-like ways to organize and regulate the conversation with the learner (Cassell 2000a).[68]

Whereas cuteness and entertainment effects soon fade away, it is the added value of a pedagogical agent's ability to communicate that makes learners *want* to take advantage of the agent's assistance regularly and over a longer period of time (Bickmore 2003; Bickmore & Picard 2005). This willingness of learners to make the pedagogical agent a part of their daily routine is essential if the agent is to enjoy lasting success (Gulz & Haake 2006: 335).

[67] The design principles discussed in this chapter were previously published in Franke (2006: 232–245).

[68] The verbal and non-verbal conversational capabilities of embodied interactive (pedagogical) software agents are discussed in detail in Chapter 10.1 and Chapter 11.

7.2 Perceptible Qualities

It is highly controversial whether computer-based artifacts ever will (or should) possess intelligence, emotions, personality, and other qualities normally attributed to humans. People have embraced or abhorred the idea of intelligent, affective artificial entities and their practical and philosophical implications, and affirmed or denied the possibility of building them. The present work will not be able to settle this issue, but it should be pointed out that when people interact with agents, it becomes more important whether an agent *seems* emphatic, intelligent, and life-like (Hayes-Roth 2004: 452) than whether it actually *has* emotions or human-level intelligence (cf. Chapter 6.5). Agents' *possession* of these qualities is less important than users' *perception* of them in an agent (Isbister 1995). Tognazzini (1993) made essentially the same point, comparing Laurel's (1991) split between the invisible computer and the visible character in the interface to the distinction between a magician and the role he or she plays on stage:

> *A magician is an actor playing the part of a magician. (...) [T]he magician is not supernatural; the character he plays is. The computer is not capable of human intelligence and warmth; the character we create is. People will not end up feeling deceived and used when they discover, as they must ultimately, that the computer is nothing but a very fast idiot. (Tognazzini 1993).*

It follows that learners' perception of the qualities of a pedagogical agent is critical for its success. If the agent exhibits cues[69] that create and maintain the illusion of life, intelligence, and interest in the learner's success and failures, it will be able to attract learners and keep them interested. Hence, pedagogical agents should be designed "through the eyes (and ears) of the learner" to make sure that learners interacting with an agent can suspend their disbelief (Hayes-Roth 2004: 461) and perceive the agent as friendly, helpful, and intelligent.

Two different qualities compete in the learner's perception of an agent: its perceived intelligence ("traditional IQ") and its perceived likeability ("social IQ"). An agent that shows off its intelligence may be less likeable ("just another brainiac") while an agent with too much appeal may be perceived as less intelligent ("just another pretty face"). Both qualities are necessary, but they may have to be traded. As Reeves and Nass put it:

> *Intelligence is a holy grail. But this may be short-sighted. When you think about the teachers that you value the most, are they the ones who were most intelligent, as in raw intellectual horsepower, or those who knew enough, and were nice – those with high social as well as traditional IQs? (Reeves & Nass 1998: 72).*

While increasing likeability may be of interest to developers of technology in general, designers of pedagogical agents should go beyond attempting to create a likeable agent and focus on aspects such as communicative ability (cf. Chapter 11), expertise (cf. Chapter 12), and emotional intelligence (cf. Chapter 13.4), which contribute to the agent's effectiveness in its supportive role with respect to the learner.

Another trade-off is between appearance and competence. In general, agents can go quite far by appearing knowledgeable without being overly competent. Humans assess performance based on simple rules. Designers can exploit these rules to build agents that appear more sophisticated than they actually are (Hayes-Roth 2004: 453f.; Nass & Brave 2005: 152).

[69] Isbister (1995) proposed a number of dimensions for assessing the intelligence of an interface agent (cf. Chapter 3.2.5), including the agent's physical characteristics (i.e. its appearance, cf. Chapter 9.2), behaviors (cf. Chapter 10), language (cf. Chapter 11), and social skills (cf. Chapter 10.3).

Creating an attractive face and voice for an agent, labeling the agent as a specialist, reminding the user that his or her success depends on the agent, conforming to gender stereotypes in appearance (cf. Chapter 9.2) and behavior (cf. Chapter 10), having the agent flatter the user, and designing a personality for the agent that matches the one of the user are all ways to increase the perceived competence of agents and other kinds of interfaces (Nass & Brave 2005: 152, cf. Chapter 6.5). However, deception, while arguably necessary to some extent when dealing socially with other individuals, can be a dangerous strategy for agents in the domain of education. Depending on their role, pedagogical agents differ from other types of agents in the amount of knowledge about the subject matter they need in order to support instructional interactions (cf. Chapter 12.3). An agent that appears knowledgeable but is unable to answer students' questions or fails to understand what the learner is doing will quickly destroy the impression of competence. Thus, the brain of an agent for education (cf. Chapter 12) must have capabilities that correspond more closely to the agent's appearance than would be required of an agent for entertainment (Johnson et al. 2000: 48). As summarized by Masterton:

> *A common problem with AI programs that interact with humans is that they must present themselves in a way that reflects their ability. Where there is a conflict between the ability of the system and the users' perception of that ability a breakdown occurs and users may either fail to exploit its full potential or become frustrated with its shortcomings. (Masterton 1998: 215).*

7.3 Balanced Design

One recurring pitfall when designing pedagogical and other agents is to make some of their aspects (e.g. character animation, emotion modeling, synthetic speech, etc.) very sophisticated (because the development team's main research and/or expertise is in these areas or some high-quality third-party components can be obtained) while others are not present or only rudimentarily developed. The result is an *unbalanced* system, which suggests human-level sophistication in some areas but fails in others when users put it to the test because they expect the agent to perform like a human in those areas as well, based on an incorrect mental model of the system (Norman 1997; Dehn & van Mulken 2000: 2; Wilson 1997, cf. Chapter 6.3) that was induced by the human-like features of the unbalanced design.

Nass and Brave (2005) used the term *mindless maximization* for the tendency of developers to realize every aspect of a user interface with the most advanced technology available, even if it means that some dimensions of the interface will be clearly superior to others and hence the overall interface seem unbalanced. They noted that this strategy, while likely to elicit positive responses from marketers, colleagues, reviewers, and award committees, is often not well received by users, causing them to feel mistrust, frustration, and dislike with respect to the interface. Nass and Brave discussed two problems with mindless maximization that explain these negative responses (Nass & Brave 2005: 150–153):

1. *Contrast highlights failure*. The human brain accentuates the contrast between the better and the less sophisticated parts of an interface, labeling the former as very high in quality and the latter as very low in quality. Two cognitive biases prevent perceptions of very high quality and very low quality from balancing each other out. First, negative experiences tend to be more arousing, memorable, and noticeable than positive ones, so in users' perception, the flaws of an interface will be more important than its features. Second, perceived poor quality in individual components generally makes the interface as a whole appear less predictable. Together, these two biases induce more negative overall judgments of interfaces of mixed quality.

2. *Smart interfaces can often seem dumber*. Even the most sophisticated technology can make an interface seem dumb if its performance cannot compete with that of the average human being, which is still largely the case when the interface involves technologies for processing speech, language, or emotion (cf. Chapter 11.1 and Chapter 13.4). For example, an interface that generates natural language output may be perceived as less intelligent because it fails to produce messages that are appropriately worded and stylistically consistent (cf. Chapter 11.1.4). Likewise, a computer that can recognize the user's speech may be penalized for high error rates. Finally, users are likely to criticize a machine which delivers messages with both positive and negative content in a cheerful voice for its lack of emotional intelligence.

The general lesson for the design of agents and other artifacts is that designers should rein themselves in sometimes when they could do better, and generally attempt to keep all aspects of an agent (or artifact) in approximate balance with respect to quality (Nass & Brave 2005: 151). The design should not raise expectations which the agent cannot meet since this will lead to frustration and the other kinds of negative user reactions mentioned above:

> *Cheating becomes visible in situations when the complexity of the agent's appearance does not match its behavioral and interactive potential. An unbalanced design might result in a mismatch of users' expectations and agent's performance, resulting in a rising frustration level of the user. (Dautenhahn 1999).*

In a balanced design, there is no single feature or capability of an agent that stands out at the expense of others. The performance of all components is on the same level. For example, the agent should not deliver carefully worded messages using a naturally sounding voice (cf. Chapter 10.1.1) if it is unable to understand a single word that the user says. Neither should it display a charismatic character without an interesting back-story to support it.

Nass and Brave (2005) argued that designers and marketers, rather than blindly picking the most advanced technologies available all the time, should base such decisions on how users *perceive* and *respond to* (different versions of) the interface. The constant desire to offer more and more advanced features, while undeniably necessary for progress, can result in unbalanced and hence less usable systems because human-level performance of all components is not feasible in the near term. Nass and Brave called for *mindful manifestation* rather than mindless maximization in interface design. That is, designers should let users be the judges of interfaces being created, carefully considering for each aspect of an interface how it interacts with and influences users' attitudes, behaviors, and cognitive processes (Nass & Brave 2005: 155).

7.4 Coherence

While interacting with the learner, a pedagogical agent must convincingly play its part in the evolving instructional conversation taking place in a dynamic learning environment. All its behaviors (speech, locomotion, gaze, gestures, postures, and facial expressions, cf. Chapter 10.1) have to be designed for *coherence*, i.e. they must fit seamlessly together and into the ongoing flow of the interaction (Hayes-Roth 2004: 460). More specifically, agents should be designed to display (p. 460):

- *Strong local coherence*. Within an episode of the ongoing interaction, agents have an immediate (though short-lived) local focus of attention and coordinate all aspects of their verbal and non-verbal communicative behavior in order to serve that focus. Clear signals mark the transition between episodes. For example, a pedagogical tutor agent may take the blame for and express its distress with the user's failure on a test for which it has prepared the learner and then make the transition to relaying the good news of the learner having been admitted to an important course in the curriculum.

- *Moderate global coherence*. Across a series of episodes and interactions, the agent coordinates persistent elements of its conversation, gesture, facial expression, and body language to maintain an enduring persona. For example, an agent for conversation training portraying an English lord will always be the same kind of snobbish gentleman with a monocle and an elegant dress, who likes to have his tea served at the same hour every day. However, agents should also make use of variable context-dependent behavioral elements to create the impression of progression within and between episodes. For example, a tutor agent will be careful not to repeat itself in its interactions with a given learner. Furthermore, its intimacy with the learner will increase over time through the exchange of personal information.

Glitches in behavior are not acceptable because they diminish the believability of the agent. In environments that change frequently and unpredictably, including interactive learning environments, coherent behavior cannot be achieved by relying (completely) on pre-planned behavioral sequences but must be generated dynamically in order to enable the agent to handle unexpected events and interruptions in a flexible way and then resume the original interaction at the point where it left off. To behave coherently and keep its overall focus in a dynamic environment, the agent needs to maintain a rich representation of context (Johnson et al. 2000: 67f.), which includes the following aspects:

- *Pedagogical context.* The instructional goals and a model of the learner's knowledge;
- *Task context.* The goals of the task in which the agent and the learner are currently involved, the current state of the learning environment, and necessary steps toward task completion;
- *Dialogue context.* A model of the ongoing conversation between agent and learner;
- *Social context.* The current state of and directions for the agent-learner relationship.

7.5 Consistency

People bring their own sets of expectations to their interactions with artifacts, especially about how the artifacts could be used. Well-designed artifacts provide intuitive affordances (cf. Chapter 6.1) that meet users' expectations about how to interface with them.

Embodied interactive agents can take advantage of natural human social affordances (Haake & Gulz 2007: 13). People know how to assess and communicate with other people, and they apply the same criteria to anthropomorphic agents. In particular, they prefer and rely on *consistency* when they interact with people and agents. That is, appearance, attitudes, behaviors, and language should all convey the same general impression of an individual because people use all the cues available to form an overall opinion (Reeves & Nass 1998; Nass et al. 2000a). Consistency not only facilitates information processing and allows the transfer of knowledge across situations, individuals, and artifacts, but also strongly influences user attitudes and emotional states due to the importance which people place on conformity to social attributes and norms (Nass et al. 2006: 374). Hence, designers of characters for cartoons, movies or intelligent agents should take great care to make their characters consistent, i.e. free from contradictions within and among their appearance, behaviors, and capabilities.

On the other hand, it has been shown that simultaneous conflicting cues from an on-screen character will not only be noticed but also disliked by people (Nass et al. 2000a). For example, it would be fatal if an agent was suggesting one type of personality with its voice and a completely different one with the way it looks or moves. Human observers would be confused because they normally rely on consistency in others to predict their actions and personality. Nass and Brave (2005) warned against confusing users with respect to the gender of voices, arguing that:

> *[P]eople strongly prefer clarity in classifying and categorizing people. Anyone or anything with an ambiguous voice is classified as strange, dislikable, dishonest, and unintelligent. Therefore, whether using a recorded or a computer-synthesized voice, clarity of gender (and other characteristics) is critical. (Nass & Brave 2005: 16).*

Inconsistent cues perceived from an agent may undermine the agent's relationship with the user, diminishing both the agent's likeability and its influence.[70] Therefore, especially designers of agents intended to serve as tutors, guides, or salespeople should pay particular attention to consistency (Isbister & Nass 2000: 264) and make every aspect of their agent consistent with every other aspect. Since agent design is commonly a team effort (cf. Section 7.12), a high level of coordination is required throughout the design process to ensure consistency.

7.6 Completeness

One of the major challenges in developing agents is to make them 'complete,' i.e. to integrate all relevant components into a coherent (cf. Section 7.4) and consistent (cf. Section 7.5) whole. A complete embodied interactive agent has:

- A 'face' to communicate emotion and conversational functions;
- A 'body' to act in the environment and in conversations with the user;
- 'Ears' and 'eyes' to perceive users and their actions;
- A 'brain' to be able to understand and (re)act intelligently;
- A 'tongue' to communicate messages verbally;
- A 'heart' to understand how the user feels and to respond appropriately;
- A 'life' to give the agent's personality depth and credibility.

Completeness is essential for the agent to make a positive impression on the learner (cf. Section 7.2), whereas incompleteness would undermine the agent's position. For example, an embodied agent that moves its lips but does not produce audible speech would be incomplete from the user's point of view, as would a pedagogical agent that appears life-like but is unable to answer questions or give explanations. A similar case for the completeness of animated pedagogical agents was made by Gulz and Haake:

> *The impact and engagingness of an animated pedagogical agent ultimately depends upon the agent as a whole – on a gestalt phenomenon including all visual aspects, together with other aspects, such as voice, dialogue, communicative style, facial expressions, and the design of the*

---

[70] For example, Gong and Nass found in two studies (Nass & Gong 1999; Gong & Nass 2000) that adding a *synthetic* face increased the social appeal of a *synthetic-speech* interface (face and speech consistent) but decreased the social appeal of a *recorded-human-speech* interface (face and speech inconsistent). From their results, they recommended that face and speech should be kept consistent in an interface, i.e. a synthetic face should be used with synthetic rather than recorded human speech.

> *underlying intelligent tutoring system. The whole is more than the sum of its parts. (Gulz & Haake 2006: 329).*

Intelligent agents bring back onto the agenda the problem of developing a complete artificial intelligent entity, which dominated the research on artificial intelligence (AI) in its early days (Hayes-Roth & Doyle 1998: 202; Sengers 1998: 3f.). Most work in AI during the past decades has focused on solving specific subproblems of general intelligence, such as machine vision or playing chess. An agent, in contrast, requires much broader capabilities, which integrate solutions from previous research on specific aspects.

Completeness should not be confused with perfection (in the sense of coming close to human capabilities). For many tasks, human-level performance is not required. What matters, however, is that all parts of the agent fit seamlessly together and perform on the same level (cf. Section 7.3). Furthermore, it should be stressed that completeness of design includes all *relevant* aspects of an agent but not necessarily the entire list given above. For instance, some pedagogical agents may have administrative tasks (such as reminding the student of important deadlines), which do not require a body or the ability to speak, so those aspects would not have to be part of their design.

7.7 Comprehensibility

People have a preference for dealing with persons and artifacts that they can understand, based on their observed actions and qualities. Hence, for many applications, it is not enough for agents to be effective because users try to discover not only *what* the agent is doing, but also *why*. Furthermore, they do not see events in isolation but try to interpret them in relation to other events (Sengers 1999).

Hence, agents should be *comprehensible* to people. That is, they should display behavioral cues that enable users to accurately interpret their beliefs, knowledge, personality, etc. Maximally comprehensible agents clearly express to the user what they are doing and why, and they make the relationships between their actions transparent to the user and how the actions fit together into a logical and coherent whole (Sengers 1999). This includes ensuring that the user understands the diachronic structure of an agent's behavior, i.e. how its actions relate to each other over time (Sengers 2000).

Comprehensible behaviors have to be simple enough that agents can communicate them adequately. Sengers proposed the following heuristic for designing comprehensible agents:

> *Behaviors should be* ***as simple as possible****. The agent's comprehensibility comes from thinking out the* ***connections*** *between behaviors and* ***displaying*** *them to the user. (Sengers 1999, her emphasis).*

Along similar lines, Wexelblat and Maes (1997) discussed the need for interface agents (cf. Chapter 3.2.5) to exhibit an "understandable competence," which enables users to understand the capabilities of an agent. In the first place, users have to know what actions of the agent produce what results. In addition, information about how operations are performed can promote users' trust and sense of control.

How do humans try to understand the actions of agents? The view of *narrative psychology* (Bruner 1986; Sarbin 1986; Bruner 1990; Bruner 1991; Crossley 2000) is that humans, being "narrative animals" (Mateas & Sengers 1999), make sense of their experience by constructing and telling stories, or *narratives*, and by listening to and interpreting the stories created and communicated by others. The human capacity to organize experience into narrative form is called *narrative intelligence* (Blair & Meyer 1997; Mateas & Sengers 1999; Sengers 2000). This line of research regards narrative as fundamental to human understanding of the behavior of humans and other intentional beings:[71] people interpret intentional actions of other individuals by assimilating them into narrative structures (Bruner 1990; Bruner 1991). Building these structures requires sophisticated interpretation skills that, for example, determine how the observable actions of others over time relate to each other, speculate about what others think or how they feel about what they do, and interpret the behavior of others in their physical, social, and behavioral context (Sengers 2000).

The implication for agent design is that an agent can become more *intentionally comprehensible* (Sengers 2000) to users if it allows people to use their well-developed narrative skills in interpreting the agent's actions, by providing narrative cues, i.e. communicating and otherwise performing in ways that users can easily[72] organize into a narrative explanation of the agent's behavior (Don 1990; Sengers 1999; Sengers 2000). Comprehensible narrative agents carefully choose and manage behavioral cues so that users can construct their own story of an agent's activity and thus create their unique understanding of this experience. In building narrative, people organize and interpret the narrative with respect to their own body of lived experience, including the things they have done and that have happened to them, the people they have met and lived with, their attitudes, beliefs, and expectations, and so on. Therefore, designing an agent to communicate a single default narrative is a one-size-fits-all approach that is unlikely to work with most users. Narrative is constructed from bits and pieces rather than absorbed in finished form, so agents should not provide pre-packaged stories of their activities but rather cues that can serve as building blocks out of which the individual user can build his or her own understanding (Sengers 2000).

Beyond helping users to understand an agent's actions, narratively comprehensible behavior also supports the user in making sense of the agent itself as an intentional being with its own attitudes, desires, and feelings (Sengers 2000). In addition, there are further possibilities to integrate agents and narrative: agents can act as characters in interactive fiction (Aylett 1999), they can tell stories, for example as part of their job as a tour guide (Isbister & Doyle 1999) or to build rapport and credibility with the user (Bickmore & Cassell 1999), and they can use narrative like humans do to understand the user and themselves.

7.8 Individuality

People stereotypically expect computers to be reliable (performing consistently without failure) and predictable (leaving no doubt about how they will behave in a particular situation). In

[71] *Intentional beings* are (perceived as) living, sentient creatures with their own thoughts, feelings, beliefs, and desires, which drive their behavior.

[72] The path of least effort would be for the agent to rely on stereotyped actions: always doing the same thing for the same reasons in the same way, thus achieving maximum predictability and reliability. However, the resulting narrative (and hence the agent) would quickly bore the user. Therefore, while agent designs should provide sufficient cues for the user to construct a narrative, they should also include elements of idiosyncrasy and unpredictability in order to give the user some work to do in narrative creation (Sengers 2000, cf. Section 7.8). Complete unpredictability or lack of convention is not advisable either, tough. An agent should appear familiar enough to users to be recognized as being similar to themselves while going against expectations enough to become and remain interesting (cf. Chapter 16.3.5.3).

contrast, they would be surprised and possibly even annoyed by an extreme level of reliability and predictability in other humans. In fact, the small inconsistencies in human behavior and the idiosyncrasies[73] of individuals (even quirks) are what make them attractive to others (Hayes-Roth & Doyle 1998: 208). Furthermore, individuals stand out because of their unique back-stories, personalities, and emotional dynamics (Hayes-Roth 2004: 455).

The implication for designing embodied interactive agents intended to be believable as an individual is to make them not overly reliable and predictable because this would diminish or perhaps even destroy the illusion of life. To manifest their own unique persona, agents should be designed with emotions (cf. Chapter 13.1 to Chapter 13.4), colorful personalities (cf. Chapter 13.5), interesting back-stories (cf. Chapter 13.6), idiosyncrasies, and (where appropriate) signature behaviors[74] (cf. Chapter 10.2) and reveal and express their personal (as an individual) and social (as a member of a group) identity (De Angeli et al. 2001b) in every aspect of their beings and behavior (Hayes-Roth 2004: 455). Back-stories should be told bit by bit in conversations with users who try to get to know a particular agent better. Reciprocal sharing of more and more personal details contributes to trust, closeness, and liking (Bickmore & Picard 2005), and is therefore an important strategy for relationship building (cf. Section 7.16).

However, it is not advisable to overdo individuality. A pedagogical agent that is more concerned with presenting itself than with facilitating learning is obviously of little use to learners. In general, an agent should always be aware of and stay true to its supportive role (cf. Section 7.14) and never draw more attention to itself than appropriate.

7.9 Variability

Few things are more disruptive to believability than repetition. In contrast to machines, people never do the same thing twice in exactly the same way. Hence, the illusion of life is quickly destroyed if an agent always delivers a given message using the same wording, tone of voice, and facial expression; follows the same steps in the same order when it performs an action; gives ‘canned’ replies to questions, no matter what went on earlier in the conversation; or displays a single behavior in response to a whole class of events.

Variability of behavior becomes more critical with increasing frequency, recency, or importance of the agent’s message and the length or memorability of the human-agent dialogue (Hayes-Roth 2004: 459). These characteristics apply to instructional interactions between learners and pedagogical agents that serve as regular companions or tutors of the learner over longer periods of time.

A believable agent must be able to vary the what, when, and how of its behavior in a way that seems incidental and unconscious (Hayes-Roth 2004: 459) and is perceived as natural and normal (Hayes-Roth & Doyle 1998: 208), not doing the same thing twice in (exactly) the same way to avoid the impression of repetitive behavior. “The greater the number of alternate ways a character has to perform a particular behavior, the stronger and deeper the perceived illusion of life” (Stern 2003: 345). Repetition is only acceptable to create an intended effect (e.g. in signature behaviors) (Hayes-Roth 2004: 459).

---

[73] *Idiosyncrasies* are peculiar patterns of behavior or thinking that set an individual apart from others.

[74] *Signature behaviors* are reliably and predictably performed by a character to establish its unique identity. For example, the cartoon character Bugs Bunny ([INT 50]) is famous for the signature greeting “What’s up, Doc?”

7.10 Communicative Ability

Agents that are designed as animated screen characters with a body, a face, and a voice invite conversational interaction (Laurel 1990b: 359, cf. Chapter 11) because they are perceived as providing similar affordances for face-to-face communication as people are familiar with from human conversation partners (cf. Section 7.5). Nass et al. found in various empirical studies that humans show social responses not only to real people but also to mediated representations, human or otherwise (Reeves & Nass 1998; Nass et al. 2000a, cf. Chapter 6.5). People respond to the representation itself rather than to the author or programmer (Reeves & Nass 1998: 254). Research focused on giving machines natural language (face-to-face) interaction capabilities is based on the common assumption that users, being experts at social interaction, prefer to interact with machines like they do with other people (Fong et al. 2003: 145f.). There seems to be some preliminary evidence for this view (e.g. Qvarfordt et al. 2003; Veletsianos et al. 2005; Nguyen et al. 2007). Therefore, embodied pedagogical agents should be able to take advantage of all available communication modes (language, gestures, actions, facial expressions, etc.) in the exchange of messages with the learner (Hayes-Roth 2004: 452). Like people, they should be able to engage in mixed-initiative dialogues with learners, where both parties can initiate exchanges and respond to contributions from the other party (cf. Chapter 11.3.3.2.2). Agents need to be enabled to perceive their human conversation partners in the same ways as people perceive each other in face-to-face conversation, and to convey the same verbal and non-verbal cues as their human counterparts would expect from another human in this context. All this amounts to designing embodied interactive agents for mixed-initiative multimodal interaction (cf. Chapter 11.3.4).

However, especially web-based environments (like the Virtual Linguistics Campus, cf. Chapter 2.4) are characterized by an asymmetry in the sophistication and availability of input and output modes. Today, an animated web-based agent can produce speech, locomotion, gaze, gestures, postures, and facial expressions while users can only express themselves using keyboard and mouse. Many users still do not have input devices like cameras or microphones, which would be necessary to capture their side of a spoken face-to-face interaction.[75] The other problem is that the technologies for recognizing and understanding speech, gestures, facial expressions, etc. are improving, but their performance is often not good enough yet for real-world applications (cf. Chapter 11.1 and Chapter 11.2).

The restricted input channels limit the amount and diversity of information that a web-based pedagogical agent can obtain about the learner, which gives rise to uncertainties in the interaction, for example about how to interpret prolonged periods of learner inactivity. For the learner, interaction through keyboard and mouse is much less natural than face-to-face interaction with other people. Therefore, the design and implementation of interactive agents for education and other applications on the World Wide Web should focus more on developing the range of input modes that can be processed by an agent, allowing conversations to move from typing to speech and from point-and-click to gesturing. The challenge is to:

[75] Another question is whether all users would tolerate the intrusion into their private space resulting from using a computer with 'ears' and 'eyes.' At least some might be reluctant to reveal auditory or visual information about themselves and their surroundings in online interactions and rather prefer to remain 'unheard' and 'unseen.'

- *Integrate and synchronize incoming information* from multiple input channels in order to understand how the meanings contributed by each of the communication modes combine into the actual message intended by the user (cf. Chapter 11.2);
- *Select and coordinate media and behaviors* (i.e. decide what media and behaviors to use, what information to express using each medium/behavior, and how to combine media and behaviors) in order to produce a coherent overall response that is tailored to users' abilities, preferences, and tasks;
- *Develop and improve component technologies* (e.g. gesture recognition, face tracking, natural language understanding, speech synthesis, etc.).

Communication not only occurs between agents and users but also between agents. For example, in virtual team-training situations involving both human and artificial team members, pedagogical agents have to coordinate their behavior with human and agent teammates (Rickel & Johnson 2003). Among themselves, artificial agents typically communicate using a formal agent communication language (cf. Chapter 14.1.5).

7.11 Modularity

The development of complex interactive software systems in general and of embodied interactive agents in particular is a time-consuming and cost-intensive process requiring many people doing many months of design, implementation, and evaluation. To reduce the costs of future projects, both agents and their components (e.g. character animation or speech synthesis) should be designed for reuse in other domains or in other agents. This can be achieved by designing the agent as a system consisting of modules that are integrated into an overall architecture.[76] One set of modules provides the agent's core functionality (e.g. user modeling, language processing, face recognition, etc.), which is domain-independent, i.e. remains stable across applications. Another set comprises the resources required to adapt the agent to a new domain, in particular the domain model (cf. Chapter 12.1). Each module should be designed as generally as possible and provide well-defined interfaces to other modules. Principles of software engineering should be observed in the development of the modules (Ghezzi et al. 2002).

Not only can a team of agent developers benefit from the ability to reuse their own modules at a later point or in another context, but it can also reuse the work previously done by others to save time and money in the development process. In fact, a growing number of technologies and resources for implementing the components of embodied interactive software agents can be purchased as commercial products (e.g. speech recognizers and text-to-speech systems); are made available free of charge (e.g. agent frameworks, character players, and dialogue toolkits), possibly including their source code; or have been built by (members of) the developer team itself in an earlier project.

---

[76] While modularity is a proven technique to facilitate software development, Sengers (2000) opposed using this 'plug-and-play' approach in agent design from a narrative intelligence perspective (cf. Section 7.7), arguing that it undermines users' perception of an agent as an intentional being when the separately designed components of the agent are visibly disjoint in its behavior. This is particularly likely to happen when the modules are designed by several developers from different backgrounds and locations.

As a general guideline, available technologies and resources should be preferred over proprietary developments wherever possible, i.e. where they significantly reduce the effort of the adopting team of developers. This indicates that a less than optimal choice of a third-party product might actually do more harm than good. Hence, the following questions may be worth considering during make-or-buy decisions:

- *How much may an external product cost?* Free software may be inexpensive but also less sophisticated than commercial packages. On the other hand, for some problems to be solved in building embodied interactive agents, free solutions may be all that is available because the technology is not yet mature enough or has not yet generated enough interest to find its way into commercial products.
- *How much effort is required to customize and integrate the product?* Any available software is of little value if it takes months to make it work.
- *Does the chosen product have the same level of sophistication as the other parts?* Choosing the 'best' option might lead to an unbalanced system (cf. Section 7.3) if the other components cannot compete with it.
- *Does the provider of the product guarantee support and regular updates?* Abandoned technologies should only be considered as a last resort and preferably should come with their source code.

To facilitate the integration of the various modules comprising the implementation of an agent, a general architectural framework that easily fits different applications would be welcome; however, such a reference architecture for agent development is not available yet. Ideally, a reference architecture for embodied interactive agents would:

- Allow rapid development of new agents;
- Facilitate testing and refining of component technologies;
- Encourage resource sharing among projects;
- Foster plug-and-play interoperability of components from different groups.

7.12 Teamwork

To create successful embodied interactive software agents for any application, it is necessary to draw upon everything that is known about the human nature and condition, both individual and collective, and combine it with our knowledge about the nature and workings of machines. That is why agent development has to be *multidisciplinary*, with input ideally coming from the entire spectrum of academic and commercial efforts investigating the different aspects of being human and (or vs.) being a machine.

The design of embodied interactive pedagogical software agents is thus both an art and a craft, involving artistic as well as scientific and technical considerations, and can only be accomplished by a multidisciplinary team bringing together human resources with multiple backgrounds and talents (Lester et al. 1997b),[77] including the following:

- *Graphic artists and animators* design the agent's appearance and behaviors.
- *Audio engineers and musicians* create the agent's musical and sound effects.

[77] Interdisciplinary teamwork is also a characteristic of e-learning development in general.

- *Voice actors* deliver the agent's verbal messages. Alternatively, the agent can use synthetic speech, which increases the amount and flexibility of its messages (cf. Chapter 10.1.1).
- *Actors* model facial expressions, gestures, and body postures and movements for the agent.
- *Programmers* develop, adapt, and integrate new and off-the-shelf agent modules.
- *Teaching experts* contribute expertise on teaching and learning methods.
- *Psychologists* supply models of emotion, personality, and user behavior.
- *Domain experts* provide knowledge about the agent's subject domain.
- *Human language technologies experts* create the speech and language processing modules of the agent.
- *Knowledge engineers* build the agent's knowledge base and reasoning capabilities.
- *Human-computer interaction experts* contribute insights into the design, implementation, and use of human-computer interfaces.
- *Evaluation experts* guide agent development through formative (process-oriented) and summative (outcome-oriented) evaluations (cf. Chapter 15).

It should be noted that this list represents an ideal assembly of human talents for designing and building an embodied interactive agent. In practice, however, not all of these talents may be present. Furthermore, multiple talents (and the associated responsibilities) may reside in one team member, due to lack of people with the required expertise, or restricted funding.

7.13 Participatory Design

Pedagogical agents are usually designed for a group of learners, which may be more or less clearly defined. Some of the attributes that characterize the target group include age, gender, ethnicity, social class, education, occupation, and organization (cf. Chapter 6.1). The composition of the target group affects the design of a pedagogical agent. The design is facilitated if the target group is homogeneous, consisting of members with the same or very similar characteristics. In contrast, heterogeneous groups, which are often encountered in web-based e-learning environments, can make design difficult because cultural and individual differences need to be reconciled. For example, an agent will have to perform an apology in different ways for students from Japan and the United States, so several variants of the same behavior have to be implemented.

It is highly advisable that designers get to know the learners that 'their' pedagogical agent will be dealing with. Otherwise, the agent may fail to meet the expectations of the target group. The best way to achieve this is through early and continuous involvement of learners, instructors, administrative staff, etc. in the design process, i.e. by designing the agent *with* them rather than *for* them. *Participatory design* (Muller & Kuhn 1993; Schuler & Namioka 1993; Muller et al. 1997) attempts to actively involve the end users of an artifact in the design process to help to ensure that the outcome meets their needs and is usable (cf. Chapter 6.3). One approach to participatory design of agents is to have users from the target group interact with a prototype agent and to include their feedback in a refined version with which the whole cycle is repeated. It can also be beneficial to make membership in, or at least familiarity with, the target audience a criterion for selecting the members of the team designing the agent because "designers do best when they create from and for what they deeply know" (Isbister 2006: 46).

7.14 Role Awareness

All the world's a stage,
And all the men and women merely players:
They all have their exits and entrances;
And one man in his time plays many parts.
(William Shakespeare. As You Like It. Act II, Scene 7.).

Pedagogical agents are more than just another technology used in instruction. By their nature, they are much closer to the human participants in e-learning experiences than ordinary computers, being *social actors* (Fogg 2003) that play various instructional roles (cf. Chapter 16) in learning environments. A given instructional role affects all aspects of a pedagogical agent, including appearance, attitudes, behaviors, capabilities, language style, and tasks.

Designing an agent for a specific instructional role helps to improve its performance, as it constrains the range of actions learners will take in their corresponding social roles (Isbister & Hayes-Roth 1998). In general,

> *role-appropriate behavior by animate characters [(or agents)] can have a powerful effect on shaping the behavior and experiences of the people with whom they interact. (Hayes-Roth & Doyle 1998: 207).*

By knowing the social role of another person, people are able to determine what the person is like and appropriately modify their behavior toward that person. As a result, it becomes easier for people to interact with others because the social roles involved shape expectations about how each participant will act and relate to others, without the need for closer acquaintance. Relying on social roles reduces the potential for embarrassment, confusion, or unwanted conflict in interactions, and it increases the stability of social groups since their members can expect each other to be able and willing to behave according to their social roles (Isbister 2006: 226).

By adopting a given social role, an individual puts on a social 'face' or mask that signals to others how he or she will behave toward them. People wear many different masks, depending on who they are dealing with and on the context of the interaction. Given the importance of social roles for relationships between individuals as well as larger social structures, people seek to preserve them by engaging in *face work* (Goffman 1967): they ignore inadvertent role-inappropriate behavior of others in order to help them to maintain consistent social identities, in other words, keep their social masks intact (Isbister 2006: 227). Face work is discussed in more detail in Chapter 11.3.1.

It follows that an embodied pedagogical agent should clearly portray its role in the learning environment, advertising its skills, capacities, and limitations (Dautenhahn 2004: 56f.). Ideally, the learners' mental model of an agent playing a particular instructional role, which they develop based on their role-specific expectations and their interactions with the agent, and the designers' model when they develop an agent for that role should be congruent. Both models are mediated by the agent's portrayal of the role in its interactions with the learner (cf. Chapter 6.3). Hence, the agent should play its role in a way that is consistent with and clearly communicates the design model, allowing learners to understand the portrayed role and to figure out how to engage with the agent based on its role. If the role design was informed by learners' expectations and the agent portrays the role believably and clearly, congruence of learners' and designers' models can be achieved.

When developing an agent for a particular role, it is necessary to identify what pressures and expectations from users are associated with that role and to equip the agent with the means to handle them. Agents that adequately portray established social roles may require much less sophistication in other areas (Hayes-Roth 2004: 454). For example, an agent playing the role of a coach in linguistic fieldwork (cf. Chapter 16.2.3) should be able to model skills in that domain for the learner, but it would not be expected to, say, teach them how to cook.

Roles have a cultural dimension (cf. Section 7.15) in that they reflect the underlying value systems within cultures, i.e. beliefs regulating the structure of society and the place of an individual in it (Isbister 2006: 54). Each social role is not only associated with culturally specific pressures and expectations but also with culturally defined patterns of interaction regulating how users and agents in their respective roles deal with each other (Maldonado & Hayes-Roth 2004: 167). For example, the design of pedagogical agents needs to account for the differences in teacher-student interactions across cultures (pp. 167f.).

Agent roles can consist of and may be related to other roles. A given role is assigned to one agent at a time, but it can also be vacant or occupied by different agents over time (Odell et al. 2003). Role assignment is the activity of creating a mapping between a set of (human and artificial) agents and a set of roles, making sure that each role is filled by the appropriate agent for the part and that an instructionally effective division of labor among the agents is found. At one extreme, one agent could play all the roles; at the other, each agent plays a different role. Practical instructional scenarios are somewhere between these two extremes. The classic role allocation in pedagogical agent systems involves a tutor agent and a human tutee, including all the agents presented as examples in Figure 17 (Adele, Steve, Cosmo, and AutoTutor). Other researchers have experimented with multiple pedagogical agents in different roles. One example is iSTART ("interactive Strategy Trainer for Active Reading and Thinking") (McNamara et al. 2004; O'Reilly et al. 2004), which uses an ensemble of agents playing different instructional roles (instructor, student, tutor, and tutee) that help learners to improve the depth of their reading skills by modeling and guiding the learning process through their interactions with each other and with the learner.

Conclusive empirical evidence regarding the effectiveness of a single agent playing multiple different roles in comparison to a group of agents where each agent represents a different role is still sparse. However, one experimental study (Baylor & Ebbers 2003) investigated the question whether it is more effective for learning and motivation to have one agent with combined expertise and motivational support (Mentor) or two separate agents, one with expertise (Expert) and one with motivational support (Motivator).[78] The authors found preliminary evidence for a *pedagogical agent split-persona effect* suggesting that two separate agents representing different functional roles (Expert and Motivator) may be preferable to one agent representing both of the roles (Mentor). In the study, the synergy resulting from the use of two agents playing the separate roles of Expert and Motivator produced a significant increase in learning (particularly recall), and the subjects perceived the two agents as significantly easier to learn from (Baylor & Ebbers 2003).

7.15 Cultural Awareness

The term *culture* means different things to different people.[79] The UNESCO (United Nations Educational, Scientific and Cultural Organization) defined culture as the set of distinctive

[78] On the experimental validation of these three pedagogical agent roles, see Baylor and Kim (2003a).

[79] Kroeber and Kluckhohn (1952) compiled a list of more than 200 definitions of culture.

spiritual, material, intellectual and emotional features of society or a social group. This definition includes art, literature, lifestyles, ways of living together, value systems, traditions, and beliefs ([INT 52]).

Like other artifacts, embodied interactive software agents are both part and product of a particular culture. The cultural background of an agent is usually not modeled explicitly but is taken implicitly from the culture of the agent's designers. For most current agents, including the ones shown in Figure 17, this is Western, industrialized culture (Sengers 2004).

People tend to prefer to interact with others who they perceive as culturally or ethnically similar to themselves (Maldonado & Hayes-Roth 2004: 151; Nass et al. 2000a). When people meet other individuals, they ascribe cultural backgrounds to them on the basis of *cultural stereotypes* (Maldonado & Hayes-Roth 2004: 152), i.e. simplifications about appearance, attitudes, behaviors, habits, preferences, etc. of the 'typical' members of a particular culture. For example, one stereotype about East Asians (especially Japanese) is that they often bow when meeting others. Stereotypes generally allow people to compare other individuals to prototypes stored in their minds and to make quick assumptions about someone new. This not only saves time and effort but also increases the predictability of everyday encounters because by adhering to a shared stereotypical code in appearance and behavior, people can be confident that they will make the intended impression on others (Isbister 2006: 13). More information about the psychology of stereotypes can be found in a review by Hilton and von Hippel (1996).

For embodied interactive software agents, the danger of causing misunderstanding, offense, and rejection increases in open environments like the World Wide Web, where they will usually deal with users from diverse cultural backgrounds. Adapting an embodied interactive software agent to a different culture involves careful (re-) consideration of the agent's identity, back-story, appearance, content and manner of speech, manner of gesturing, emotional dynamics, social interaction patterns, role, and role dynamics. The latter two aspects were discussed in the previous section, the others are covered in the sections below, following the account of Maldonado and Hayes-Roth (2004: 153–168).

7.15.1 Identity

Identity comprises the essence of who a particular agent is, i.e. its distinct personality as a persistent entity. This includes demographic aspects (e.g. the agent may be portrayed as a 39-year old Chinese female) and traits and qualities (e.g. likes and dislikes, signature and idiosyncratic behaviors). The concept of identity has to be distinguished from *appearance* (cf. Section 7.15.3 and Chapter 9.2). The latter is merely a graphical or other representation of certain aspects of the agent's physiological form (e.g. age, gender, race, build, hair style, etc.) and certain elements of its personality and background (e.g. dress style and social class).

Cultural norms and roles influence the identity of an agent in every aspect of the interaction. For example, if the agent is designed to portray a child, cultural conventions will regulate the appropriateness and form of addressing people of varying ages. Female agents face the challenge that the social roles of women are more variable across cultures than those of men. Furthermore, in many cultures, professions and roles are still stereotypically assigned to particular genders and agents defying a stereotype may inspire less respect or confidence (Nass et al. 1997b, but see the discussion in Chapter 9.2).

7.15.2 Back-Story

The back-story of an embodied interactive agent includes the personal experiences and history of the individual portrayed, which played a significant part in shaping the agent's identity, as well as current personal facts which illustrate that the agent has a (continuing) life beyond the screen. Back-stories may contain elements such as family relations, friendships, childhood memories, political views, favorite colors, food and sports, changing love interests, and financial status. They are influenced by the norms and values of the culture in which an individual was brought up, so designers should be careful to provide for cultural variation in the story of an agent that is localized to different cultures. The design of agents' back-stories will be discussed further in Chapter 13.6.

7.15.3 Appearance

Appearance is the first aspect of an embodied interactive agent that is perceived and interpreted by users. It influences how effectively and credibly the agent portrays its role and conducts interactions (cf. Chapter 9.2). Therefore, designers need to understand how gender roles, cultural norms, and traditional attires influence the user's experience even at first sight of the agent persona. They should choose a representation for the agent that will appeal to the target group of users and adopt a dress code that will be considered acceptable or at least neutral for the targeted population and role. Doing this, designers should keep in mind that attractive features and fashions change with both culture and period.

7.15.4 Content of Speech

Embodied interactive agents should speak in a way that is comprehensible to the target audience. Hence, the language and dialect of the agent should be as closely matched with the target culture as possible. This includes idiomatic expressions, slang, colloquialisms, and the appropriate selection of words to express different states of affairs (cf. Chapter 11.1.4).

Content of speech also concerns the choice of conversational topics. If an agent can talk about local historical events, geographical details, sports, favorite pastimes, recent local events, holidays, humor (cf. Chapter 10.3.2), etc., it will enhance its image as a representative of a particular culture.

However, there are also sensitive topics that speakers avoid discussing with complete frankness in some, most, or all forms or contexts of speech. The range and treatment of sensitive topics in conversation varies across cultures. In Western cultures, such *taboo* subjects include sex, reproduction, death, excretion, and the human body. Since taboos prohibit the use of plain language, speakers (and agents engaging in conversations with them) must find ways to avoid or handle these topics in a manner that is considered acceptable in the particular culture of their interlocutors. A common strategy is *lexical replacement*, i.e. substituting some other, socially acceptable item for a tabooed word. Another option for speakers involves *euphemism*, i.e. the replacement of a word or expression that is taboo, negative, offensive or too direct by a mild or indirect one.

Regarding the phrasing of utterances, agent design should make a difference between two major categories of cultures: individualistic and collectivistic (Ting-Toomey 1999: 66ff., cf. Table 11). *Individualistic cultures* emphasize individual identity, rights, and needs over the

Table 11. Individualistic and collectivistic cultures compared (Ting-Toomey 1999: 67, Table 3.1).

Individualistic Cultures	**Collectivistic Cultures**
• "I" identity • Individual goals • Interindividual emphasis • Voluntary reciprocity • Management of individuals	• "We" identity • Group goals • In-group emphasis • Obligatory reciprocity • Management of groups
Examples: • United States • Australia • United Kingdom • Canada • Netherlands • New Zealand • Sweden/France • Germany	Examples: • Guatemala • Ecuador • Panama • Indonesia • Pakistan • Taiwan/China • Japan • West/East African countries

identity, rights, and needs of the group (Allbeck & Badler 2004: 110f.), promoting self-efficiency, individual responsibilities, and personal autonomy (Ting-Toomey 1999: 67). Emotion, thought, and action are viewed as being motivated by attributes of the individual and as independent of his or her environment. In contrast, in *collectivistic cultures*, the identity, rights, and needs of the group take precedence over those of the individual. Collectivistic cultures value relational interdependence, in-group harmony, and in-group collaborative spirit (Ting-Toomey 1999: 67). In these cultures, social relationships, roles, and responsibilities provide the framework for organizing, experiencing, and explaining the thoughts, feelings, and behaviors of the individual.

Content of speech is also affected by the degree to which a culture prefers verbal vs. non-verbal means to communicate messages. The continuum ranges from high-context to low-context cultures (Hall 1976). In *high-context cultures*, considerable parts of a message are communicated through facial expressions, body language, and other indirect cues, rather than the actual words of the message. In contrast, *low-context cultures* tend to rely more on spoken language than non-verbal cues in communication and expect messages to be overtly clear (Isbister 2006: 51). Individualistic cultures, which value individual independence and freedom of expression, are low-context and consider communication as something that occurs explicitly through the verbal content of a message. Collectivistic cultures, which seek to preserve the harmony of the group and the social face of its members, make more use of non-verbal communication to protect the group and people's feelings (p. 51).

7.15.5 Manner of Speaking

Apart from *what* an embodied interactive agent says, further cues about the agent's cultural background are provided by *how* and *when* it delivers its messages. Intonation, pronunciation, timbre,[80] and range of vocal expression are acoustic characteristics of speech that should be varied to localize an agent for a particular culture. Other properties of utterances that convey the identity and cultural background of agents include timing, speed, utterance length, frequency and choice of gap fillers (e.g. 'hmm,' 'umm,' and 'like'), and stuttering.

Designers should pay special attention to how an agent greets its interaction partner because the first impression is the one that lasts (cf. Chapter 9.2). Agents with faces (cf. Chapter 10.1.3) are expected to be able to recognize and remember users, and to address them accordingly. Title usage and familiarity play an important role in properly addressing an interlocutor. Some cultures prefer a first-name basis even on formal occasions (e.g. Iceland and Thailand), whereas other cultures license addressing someone by their first name only after years of close contact. Titles are avoided for the most part by American speakers, but in many Latin American countries, 'doctor' is used for anyone with a degree or a position of authority.

7.15.6 Manner of Gesturing

Gestures are an important vehicle for communicating meaning in everyday face-to-face interactions. People use their face and hands in dialogue regardless of language, cultural background, or age (Cassell & Stone 1999). However, certain gestures may cause misunderstandings and even offense in cross-cultural communication because the same gesture may have different meanings for members of different cultural groups. Emblematic gestures, which symbolize something abstract, are particularly problematic in this respect. Examples of such potentially dangerous emblems include gestures of assent and dissent, such as the ring gesture and the head nod. Gestures are discussed in more detail in Chapter 10.1.2.

Gaze (cf. Chapter 10.1.4) is another example of non-verbal behavior that is associated with cultural cues and conventions. For instance, Japanese and Korean users might interpret a direct, sustained gaze of an embodied interactive agent as insulting or as a sexual advance. In contrast, users from the United States might regard an agent averting the gaze as dishonest, weak, or very shy.

7.15.7 Emotional Dynamics

The interpretation, management, and expression of emotions not only have an individual but also a sociocultural dimension. What emotion should be communicated by whom in what situation with what frequency and intensity to what addressee for what length of time, how much stimulation is necessary to trigger an emotional outburst, and how comfortable people are when directly confronted with certain emotional displays, varies from one culture to another.

[80] *Timbre* is a complex, distinctive quality of human voices which is determined by the different parts of human speech anatomy: the diaphragm, lungs, trachea, larynx, vocal cords, glottis, pharynx, oral cavity, jaw, teeth, tongue, and lips. Each of these elements, in turn, is influenced by an individual's genetics, endocrine system, health history, physical environment, and habits (Nass & Brave 2005: 98).

Agents need to deal with this multidimensional space of possibilities both when they interpret the user's emotional state and when they generate their own emotional displays (cf. Chapter 13.4.2).

Agent designers need to choose an appropriate theoretical framework that facilitates cross-cultural emotion expression and interpretation. One such framework is the *Basic Emotions Theory* (Ekman 1984; Ortony & Turner 1990; Ekman 1992a; Ekman 1992b; Ekman 1999a; Ekman 1999b). This theory is based on research by the psychologist Paul Ekman, who identified facial expressions and corresponding emotions that are not determined by culture but seem to be universal to the human species and thus possibly have a biological origin. These 'basic' emotions (anger, disgust, fear, happiness, sadness, surprise, and embarrassment) are shared and recognized by humans across cultures. Hence, they can form the basis for believable emotional displays of agents that are universally recognizable (cf. Chapter 13.4.2).

7.15.8 Social Interaction Patterns

Different cultures have different views of how and when particular topics are appropriately brought up for discussion. Hence, interactive agents should be designed to respect the rules of a culture that regulate the manner and timing of talking about business, delivering bad news, taking the next step toward familiarity with the user, and so on.

Given the present limitations of the technologies involved, agents will frequently make mistakes in their interactions with users and therefore need to be equipped with means to address their shortcomings. Socially acceptable ways for an agent to recover from a faux pas, acknowledge insufficient understanding, and apologize for a mistake depend on how mistakes are perceived in the target culture and how willing its members are to accept repeated apologies. Whereas in individualistic cultures, an agent may appropriately react defensively and non-apologetically to criticism of its performance in certain contexts, agents from collectivistic cultures may tend to accept the blame and apologize for their mistakes.

Further culturally specific concerns for the design of human-agent interactions include how often an agent should take the initiative; how and when it is appropriate to interrupt the user; the pace of turn-taking in the conversation; ways for the agent to acknowledge the contributions of its interlocutor; and the acceptable frequency of exchanging questions with the user.

7.16 Relationship Building

A *human* or *interpersonal relationship* is a particular type of association, connection, or affiliation existing between two individuals or within a larger group of people. Relationships are social and emotional constructs, which are incrementally built and maintained over a series of interactions that can span a long time (Bickmore 2003). They may vary in their level of intimacy and sharing, both across relationships and at different points in time within the same relationship. People seek relationships for the various kinds of support they provide, such as (Bickmore & Picard 2005):

- *Emotional support*, e.g. esteem, reassurance of worth, affection, attachment, and intimacy;
- *Appraisal support*, e.g. advice and guidance, information, and feedback;
- *Instrumental support*, e.g. material assistance;
- *Social network support*, e.g. providing introductions to other people.

Humans develop social-emotional relationships not only with other humans and animals but also with human-made artifacts, such as cars, houses, toys, works of art, and animated characters in film. The strength and meaningfulness of relationships between people and artifacts can be similar to those of human-human relationships (Stern 2003: 333). The nature of the relationship has a direct impact on productivity, enjoyment, engagement, trust, and other aspects of the user's experience with an artifact, and the importance of the relationship increases as the interaction extends over time (Bickmore & Picard 2005).

Recently, people have also been shown to bond with various kinds of computer-based artifacts designed for use in art or entertainment. Turkle (2003; 2005; 2006a; 2006b; Turkle et al. 2006) introduced the term *relational artifact* for "computational entities that present themselves as having states of mind that are affected by their interactions with human beings, objects designed to impress not so much through their 'smarts' as through their sociability" (Turkle 2006a). Relational artifacts are realized as animated characters or physical robots and typically invite some form of interaction with users. Many of them serve as virtual pets or companions to the user, such as Aibo ([INT 53]), Creatures (Grand et al. 1997; Grand & Cliff 1998; Cliff & Grand 1999), Dogz and Catz ([INT 54]), Furby ([INT 55]), and Tamagotchi ([INT 56]). Users take care of their virtual pets, feeding them virtual food, petting them with a virtual hand, talking to them using a microphone, and generally giving them affection and attention. Pets that are well cared for thrive, whereas those that are neglected become distressed or weaker until they run away or die. Having a virtual pet is different from playing a computer game because it is not oriented toward an end goal (typically winning the game). The enjoyable experience for the user does not come from winning but from the *process* of having a relationship with a virtual pet, and some users seem to be eager to form emotional relationships with virtual characters, showing willingness to suspend their disbelief (Stern 2003: 337f.; Evans 2004).[81] As Turkle put it:

> *When people are asked to care for a computational creature, they become attached, feel connection, and sometimes they feel a great deal more. We love what we nurture, biological or not. (...) [A]t the heart of the "holding power" of relational artifacts is that they call forth the human desire for communication and connection. (Turkle 2006a).*

Pedagogical agents are meant to be there for the individual learner, which invites a much closer social relationship than learners would form with human teachers in the classroom. This, in turn, addresses the human need for profound personal relationships with other individuals in

[81] It should be noted that owner-pet relationships are easier to establish and maintain than friendship and other more complex kinds of relationships because of their simpler dynamic and lesser need for communication between the parties involved (Fong et al. 2003). Furthermore, the relationship between a virtual pet and its owner, unlike friend-to-friend relationships, is inherently unequal (Stern 2003: 338): the user 'adopts' his or her pet (typically by buying it) and the pet is dependent on the attention of its owner. In contrast, friendship usually happens because *both* individuals concerned choose to be friends with each other, and (ideally) there is mutual give and take between equals in the relationship. This raises the question if there could ever be relationships at the level of 'true' friendship (or more) between users and human-made interactive artifacts that have a price tag attached to them (including pedagogical agents), no matter how human-like they may be portrayed. (How can something be considered one's equal if one has paid for it?) After all, the user might prefer to control rather than befriend his or her artifact (or agent), like any other acquisition. De Angeli et al. (2001a) explored users' social reactions to a chatbot (cf. Chapter 11.3.2) and found that users preferred having an asymmetric relationship with the chatbot in which they were in the dominant position.

the learning process (Gulz 2004).[82] Hence, the design of pedagogical agents should take advantage of the natural human tendency toward relationship building by making pedagogical agents *relational*, i.e. able to build long-term social-emotional relationships with their learners (Bickmore 2003). Since people build and maintain relationships primarily in the context of face-to-face conversation (Bickmore et al. 2005; Park & Catrambone 2007), pedagogical agents can implement relationship-building strategies as verbal and non-verbal conversational behavior (Bickmore & Picard 2005; Bickmore et al. 2005):

- *Social dialogue* (cf. Chapter 5.4.5 and Chapter 10.3) to bridge interpersonal distance and establish a connection to the learner;
- *Communication of empathy* (cf. Chapter 3.1.3) to provide emotional support;
- *Reciprocal deepening self-disclosure* to increase trust, closeness, and liking;
- *Meta-relational communication* ("talk about the relationship") to clarify mutual expectations at the beginning, to periodically ensure that things are running smoothly, and to make the necessary 'repairs' if they do not;
- *Use of humor* to increase liking (cf. Chapter 10.3.2);
- *Talking about the past and future together* and *reference to mutual knowledge* to indicate acquaintance;
- *Use of continuity behaviors* to bridge the time spent apart (e.g. appropriate greetings and farewells, talk about the time spent apart, etc.);
- *Emphasizing commonalities and destressing differences* to increase solidarity and rapport;
- *Mirroring or entrainment*, i.e. adoption of some aspects of the learner's behavior, e.g. vocabulary (lexical entrainment);
- *Use of non-verbal 'immediacy' behaviors* (e.g. close distance, direct body and facial orientation, and smiling) to indicate liking, conversational engagement, and solidarity (cf. Chapter 10.3.1);
- *Persistent memory* of previous interactions with the learner to provide information about the shared history for other relational behaviors (e.g. talking about the past and future together and continuity behaviors);
- *Ability to (re-) identify and distinguish between learners* to generate appropriate relational behaviors for the individual learner (Schulman et al. 2008).

7.17 Summary

The design of embodied interactive pedagogical software agents requires careful consideration of a host of artistic, engineering, and scientific aspects. Often, designers may be tempted to rely on their intuition and inspiration. However, this approach can lead to suboptimal decisions in design. As a first step toward a more solid basis for agent design, sixteen principles for the design of embodied interactive software agents for education and other domains were laid down in this chapter. These principles are summarized below:

[82] Learning environments with embedded socially interactive pedagogical agents have been called *intrinsically social* because no other human participants are required to provide the learner with social interaction, in contrast to *extrinsically social* learning environments, which rely on people to embody social agency and to enable social interaction (Gulz 2004).

- *Added value*. The agent should possess knowledge and skills (in particular in verbal and non-verbal communication) that induce users to take advantage of the agent's services regularly and over a longer period of time.
- *Perceptible qualities*. The agent should exhibit cues that allow users to perceive that the agent possesses certain desirable qualities (e.g. believability, competence, conversational ability, empathy, and personality).
- *Balanced design*. Designers should avoid mindless maximization of the sophistication of particular components of an agent and instead attempt to keep all aspects of an agent in approximate balance with respect to quality.
- *Coherence*. All behaviors of the agent (speech, gestures, facial expressions, gaze, postures, and locomotion) should fit seamlessly together and into the ongoing flow of the interaction.
- *Consistency*. Agents should not exhibit contradictions within and among their appearance, behaviors, and capabilities but convey the same general persona through all these aspects.
- *Completeness*. A complete agent should integrate all relevant components (including face, body, ears, eyes, brain, tongue, heart, and life) into a coherent and consistent whole.
- *Comprehensibility*. Users should be able to build an accurate interpretation of the agent's actions, beliefs, personality, etc. from the appearance and behavioral cues displayed by the agent.
- *Individuality*. The agent should reveal and express its personal and social identity in every aspect of its being and behavior, which includes emotional dynamics, personality, back-story, idiosyncrasies, and signature behaviors.
- *Variability*. The agent's behavior should not seem repetitive but vary incidentally and unconsciously with respect to the kind, timing, and manner of the behavior being performed.
- *Communicative ability*. The agent should be able to communicate with users and other agents. In human-agent interactions, this involves using all available communication modes (speech, gestures, facial expressions, etc.) in mixed-initiative multimodal conversations with the user.
- *Modularity*. The agent and its components should be designed in a way that facilitates their reuse in other application domains or in other agents. The design should consist of a system of domain-independent and domain-specific modules that are embedded in an overall architecture.
- *Teamwork*. Agent development is a multidisciplinary team effort involving a wide range of human talents that contribute knowledge and techniques ideally covering the entire spectrum of academic and other efforts to understand the human nature and condition.
- *Participatory design*. The agent should be designed with early and continuous involvement of its future users and employers during the development process.
- *Role awareness*. The agent should portray its social role (including its skills, capacities, and limitations) in a way that helps users to understand the role and to appropriately adapt their behavior toward the agent based on its role.
- *Cultural awareness*. The agent should be viewed as both a representative and a product of a particular culture. Its cultural background should be modeled explicitly by (re-) considering the agent's identity, back-story, appearance, content and manner of speech, manner of gesturing, emotional dynamics, social interaction patterns, role, and role dynamics.
- *Relationship building*. The agent should be made relational, i.e. equipped with strategies and behaviors conducive to building and maintaining long-term social-emotional relationships with individual users.

8 Agent Design Failures

The second half of the 1990s saw considerable interest, both academically and commercially, in the potential of software agents in general and interface agents (cf. Chapter 3.2.5) in particular. As a prominent example, the Microsoft Corporation ([INT 57]) included various embodied interactive agents in their products, which were designed as 'social' interfaces to assist users in their work with Microsoft's graphical user interface (GUI) and office application software:

- In March 1995, Microsoft Bob was released as a user-friendly interface layer running on top of the graphical desktop Microsoft Windows 3.1. Microsoft Bob was based on the room metaphor and featured several interactive assistants.
- The Office Assistant, also known as Clippy, Clippit, or the Paperclip after its default representation, was introduced two years later in Microsoft Office 97 as a social entry point to the application's help system.

Both systems were much less successful than their designers anticipated, proving neither very useful nor very popular with users. Especially the Paperclip came to be loathed by many users and was disabled much more often than its assistance was sought. Detailed instructions on how to remove (rather than use) this feature were abundantly requested and published in computer magazines and on the World Wide Web. In reaction to its unpopularity, the presence of the animated Assistant has gradually waned in successive Microsoft Office products, and in the latest version, it has even been discontinued altogether in favor of an alternative help system. Schaumburg (2001) conducted a questionnaire study in which she asked users of the Microsoft Office suite to indicate their use and acceptance of the Office Assistant. Her results confirmed users' negative perception of the Assistant, which was rejected by the majority of respondents because they thought that it was neither efficient nor engaging or motivating.

Why were these character-based systems a failure? While users often were quite emotional in their rejection of the characters, simply stating that they were 'annoying,' it appears that the reasons for the lack of acceptance are more complicated. Certainly, it would be premature to conclude from these failure (rather than success) stories that the overall concept of embodied interactive agents is misguided. However, studying these design failures can be instructive for designing more effective agents. Interestingly, there are very few scientific publications on this topic, noteworthy exceptions being Schaumburg's study and Swartz's thesis (Swartz 2003), which explored reasons why so many people responded negatively to the Office Assistant.

Coming from the discussion of agent design principles in the previous chapter, it can be argued that the failure of both Microsoft Bob and the Paperclip is due to serious flaws in their design which result from a violation of several principles of agent design. These violations are analyzed in the following two sections, first for Microsoft Bob, then for the Office Assistant. Lessons for agent design that can be drawn from this analysis are presented in Section 8.3. The chapter concludes with a summary and an outlook on the subsequent chapters on the different components of agent design.

8.1 Microsoft Bob

The Microsoft Bob environment is based on the "room metaphor." The whole interface is designed as a set of rooms within some house. Each room is furnished with images of decorative and functional objects. By clicking on the latter, the user can start different application programs, such as a word processor, a quiz, and a calendar. A screenshot of Microsoft Bob is shown in Figure 33.

The user is not 'alone' in this simulated set of rooms. First, there is a 'general' interface agent, an animated dog by the name of Rover. In addition, particular applications may have specialist interface agents associated with them. The purpose of all these agents is to assist the user in his or her work within the environment.

The following principles of agent design were observed only inadequately when designing the assistants in Microsoft Bob:

- *Added value*. It is doubtful whether the assistants in Microsoft Bob add real value to the interaction. They try hard to be helpful while appearing cute and entertaining at the same time. However, for regular or experienced users, they quickly become a source of distraction. There is no way to open an application program without having one of them advertising its help by displaying large balloons full of options. When the agent is idle, the balloon persists while the agent is performing various behaviors to enhance its believability (cf. Chapter 10.2). Although the agent is always there, users' possibilities to interact with it are fairly limited (see below). It is not possible to dismiss an agent, and the agent itself is unable to determine when its assistance is not wanted.

  It could even be argued that most applications supported by Microsoft Bob do not require the presence of an agent because they do not depend on social interaction. For example, Rover providing interactive assistance in the word processor can more hinder than help the user, who may just want to write a letter. The only application where the presence of an agent might be justified is the game GeoSafari, which features an agent hosting the quiz.
- *Perceptible qualities*. Each agent in Microsoft Bob does its best (given the limitations of the technology) to appear life-like and believable (but see the remarks about their individuality and variability below). However, it seems that the designers emphasized appearance and perceived likeability while neglecting competence and perceived intelligence. None of these characters appears overly intelligent (except for the Book-with-Glasses character in the checkbook application, perhaps), and their competence is limited to advising the user in specific application programs. Only Rover seems to possess broader knowledge of several applications, but its competence is still limited to giving advice.
- *Balanced design*. The agents in Microsoft Bob exhibit a fairly imbalanced design. The emphasis is on animation and sound, which are mainly used to make an agent more entertaining, whereas other aspects like adaptiveness and conversation are non-existent or underdeveloped. It should be noted that this also makes these agents incomplete.
- *Consistency*. There are various inconsistencies in the behavior of agents in Microsoft Bob. Rover, for example, delivers text messages in natural language but cannot speak (only bark). Furthermore, sometimes, an agent simultaneously delivers a text message and a spoken message with different contents.
- *Individuality*. Animation and sound are the chief vehicles for agents in Microsoft Bob to convey the impression of a unique persona. Each character has unique features and characteristic (though repetitive) behaviors. However, they do not possess deep and interesting personalities, although Rover tries to add some depth to its persona by telling the user about itself as part of the guided tour. In some contexts, the agents display emotion (cf. Chapter 13.4.2), but most of the time, they do not clearly portray their emotional state.
- *Variability*. The agents have a fairly small inventory of animations and sounds, and they always perform them in the same way, regardless of the current context or their internal state. Hence, there is little to no variation in their behavior. Rover actually tries to vary its

Figure 33. A room in the Microsoft Bob interface featuring several functional objects and the animated character Rover the Dog. Screenshot taken on 26 October 2007.

reactions when one clicks on its image, but after a small number of clicks, it has exhausted its repertoire and the behaviors repeat themselves.

- *Communicative ability*. Interactions between the user and the interface agents in Microsoft Bob are limited. The agents' behaviors are chiefly performed to entertain the user. They cannot speak, except for canned recorded messages in some of the characters (Rover can only bark), they rarely move, and both their task-related and their communicative gestures are sparse. They only accept keyboard and mouse input, with the latter being the prominent way to interact with an agent. This limits what the agent can learn about the user.
- *Role awareness*. In most applications within Microsoft Bob, Rover assumes the role of the assistant and portrays it fairly clearly throughout an interaction. However, was casting a dog character for this role the appropriate choice? Would users expect a dog to have competence in assisting them to write a letter, manage their calendar, etc.?
- *Relationship building*. Rover and the other agents in Microsoft Bob adopt essentially a passive strategy for relationship building. They lack explicit behaviors, protocols, and strategies for establishing, maintaining, and changing their relationship with the user. In addition, they make poor use of language, the primary means used by humans in relationship building.

8.2 The Microsoft Office Assistant

In the Microsoft Office environment, the Paperclip (cf. Figure 34) and the other Office Assistants play three major roles (Swartz 2003: 5–11):

- *Proactive help system*. The Office Assistants suggest ways in which they can help the user to accomplish a task easier or with better results.
- *Natural language help query*. The user can ask questions in (relatively unrestricted) natural language, and the Assistant uses basic statistical inference mechanisms to guess a user's goal given their request for help.
- *Locus of agenthood*. The Office Assistant serves to locate agenthood in the environment. For example, it voices dialogue alert boxes and performs actions in response to user commands (e.g. saving, printing, or sending e-mail) which suggest that the Assistant is the one carrying out those commands.

As the Paperclip and its siblings in Microsoft Office can be seen as descendants of the earlier generation of agents in Microsoft Bob, the criticism noted for the latter essentially also applies to them, in particular because their behavior and strategies of approaching (or rather disrupting) the user are quite similar, although their animations and sound effects are more advanced than those used in Microsoft Bob. Some additional points of criticism concern the following principles:

- *Added value*. The Office Assistants do not add real value to the interaction between the user and the office application. The user does not need an Assistant to obtain help because the regular help features provide the same service as the Assistant's search box. The same natural language queries can be formulated via the "Answer Wizard," with the same results. Thus, the Assistants do not have anything substantial to offer; they do not possess expertise that would make them useful.
- *Balanced design*. In contrast to the agents in Microsoft Bob, the Office Assistants have the rudimentary capability of understanding a user's typed query. While the animations and sound effects of the Office Assistants are quite sophisticated, their natural language 'understanding' capabilities are only rudimentary. This imbalance can lead users to form exaggerated expectations about an Assistant's competencies, based on their appearance and visible behaviors, which the Assistant is unable to meet.
- *Comprehensibility*. The Assistants may not be equally comprehensible to all kinds of users. For example, beginners, the people who are supposed to gain most from interacting with an Office Assistant, have been shown to be somewhat confused about what the Assistants are supposed to do. More advanced users seemed to have a clearer mental model (cf. Chapter 6.3) of the Office Assistant (Swartz 2003: 29).
- *Role awareness*. The proactive help feature seems to undermine the Assistant's supportive role. Users may perceive this feature as "challenging their authority" when interacting with the office application. They may not want the Assistant/computer to think that they need help but may rather wish to ask for it themselves.[83] A better strategy for designing help

[83] Part of the technology for the Office Assistant was derived from prototypes developed in the Lumière project at Microsoft Research (Horvitz et al. 1998). The original Lumière prototypes used decision-theoretic methods (in particular Bayesian networks) to decide when to offer help to a given user, considering the user's background, actions, and queries to infer his or her needs under uncertainty.

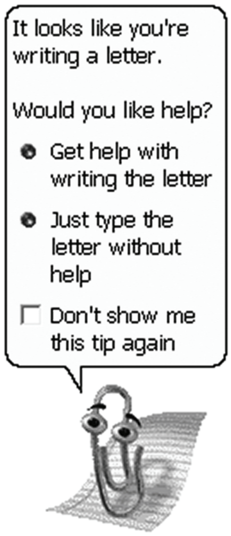

Figure 34. The Paperclip offering its infamous proactive help on writing a letter (Swartz 2003: 5, Figure 3).

agents like the Office Assistant might be to refrain from making the Assistant "hold the user's hand" all the time but to increase the user's feelings of control and self-reliance by designing the agent to teach the user a skill, empowering him or her to accomplish a task on his or her own the next time (Swartz 2003: 17).

- *Cultural awareness*. The Office Assistants ignore the sociocultural conventions regulating when and how to interrupt someone. They simply pop up uninvited, which is considered impolite and intrusive in many cultures. Furthermore, they persist in displaying their proactive letter-writing help feature, despite being dismissed by the user an arbitrary number of times, a mistake which is not easily forgiven anywhere. Then, the Office Assistants are constantly staring at the user and monitoring his or her work. Not only can this make users feel uncomfortable (cf. Chapter 10.4), but in a number of cultures, this is also considered unacceptable behavior.
- *Relationship building*. The Paperclip and the other Office Assistants adopt a passive strategy for relationship building, lacking the relevant behaviors, protocols, and strategies for a more active approach. Like the agents in Microsoft Bob, they cannot use language to form a

---

For example, assistance was only offered if the "likelihood that help is needed" by the user exceeded a given threshold, which could be raised or lowered by the user for future interventions of the system. If the user ignored the help offer, the system would step back with a brief apology (Jameson 2003: 308). Interestingly, many of the more sophisticated user-adaptive features of the Lumière prototypes were not included in the deployed Office Assistant (Horvitz et al. 1998), which may have contributed to its lack of success. In fact, one of the Lumière investigators (Eric Horvitz) wrote on his homepage:

> *The Office team has employed a relatively simple rule-based system on top of the Bayesian query analysis system to bring the agent to the foreground with a variety of tips.* ***We had been concerned upon hearing this plan that this system would be distracting to users*** *– and hoped that future versions of the Office Assistant would employ our Bayesian approach to guiding speculative assistance actions – coupled with designs we had demonstrated for employing non-modal windows that do not require dismissal when they are not used. ([INT 58], emphasis added).*

relationship with the user. For example, they are incapable of engaging the user in small talk. They cannot even answer simple questions about themselves, such as "Who are you?"

Schaumburg (2001) conducted a questionnaire study whose results indicate further reasons for the Office Assistant's lack of popularity with the majority of users. Most of the respondents in her study rejected the Office Assistant because they perceived it as 'inefficient' and "not particularly engaging or motivating." Two variables contributed significantly to the (lack of) likeability and use of the Office Assistant: users' perception of being distracted by the agent and their trust in its helpfulness, with distraction being more important than trust. "Only those users who felt distracted by the assistant and who did not trust in its capability to solve their problems rejected the Office Assistant" (Schaumburg 2001). Interestingly, while most subjects did not regard the Office Assistant as a social partner, their responses indicated that there was a social dimension to their perception of the agent: they described it as 'impolite' or 'cute' and shared that they got along with 'him'[84] and that they loved or hated "that guy." Still, as Schaumburg noted, the designers of the Office Assistant failed to take advantage of users' "social bias" in order to achieve a more positive perception of the agent.

8.3 Lessons for Agent Design

The previous two sections identified a number of shortcomings of the agents in the Microsoft Bob and Office Assistant systems with respect to the principles of agent design established in Chapter 7. These shortcomings provide several valuable lessons for designers of embodied interactive software agents, which can be summarized in the following general design guidelines:

- Do not create agents for applications to which they do not add value. In particular, if social interaction is not critical for the task, do not include an animated agent at all but look for less obtrusive and time-consuming ways to assist the user (cf. Chapter 3.2.5). The same point was made by Fogg with regard to the question when to make social cues more explicit in computing products:

  > *When I buy gas for my car, I choose a station with gas pumps that take credit cards directly. I don't want to deal with a cashier; I'm not looking for a social experience. I believe the analogy applies to interactive technologies, such as word processing programs or spreadsheets, that people use to perform a task more efficiently. For such tasks, it's best to*

[84] There are two things worth noting about how respondents referred to the Office Assistant in the comments cited in Schaumburg (2001). First, there was a tendency to apply 'animate' referring expressions to the agent (e.g. *he* rather than *it*); second, respondents used masculine rather than feminine expressions. While the former suggests that for many subjects (at least subconsciously), the Office Assistant was more than a tool or feature, compared, for example, to the spellchecker, the latter might indicate that subjects stereotypically associated the role of an expert on technology (which is portrayed by the Office Assistant) with a male individual. This hypothesis is consistent with results from Baylor et al., who found that female agents were perceived as less competent than male agents (e.g. Baylor 2005). Subjects' responses may have been influenced by the particular representation of the Office Assistant that they have experienced; on the other hand, the default character for the Assistant is a paper clip (i.e. an object rather than a person), so stereotypical expectations about expert gender might have played a role in users' choice of referring expressions for the agent.

minimize cues for social presence, as social interactions can slow things down. (Fogg 2003: 115).

However, other applications rely much more on social interaction than on efficiency and therefore may benefit from enhancing cues in computing technology that suggest the presence of a social actor. Examples include products for leisure, entertainment, and, important for this thesis, education (Fogg 2003: 115).

- Balance an agent's capabilities, so that it does not excel in some areas but disappoint in others. Do not be deceived that the agent's strong points will outshine (or outbalance) its weaknesses.
- Make the agent as likeable and as competent as required for the application, and ensure that users clearly perceive both qualities.
- Help users to form a clear mental model of the agent, which includes its identity, purpose, capabilities, actions, and state.
- Craft a character that is appropriate for its intended role and perceived by users as such, in line with the stereotypical expectations and pressures associated with the role.
- Make the agent behave according to its supportive role; the agent should empower the user, not weaken his or her sense of control and self-reliance.
- Ensure consistency within and across the agent's behaviors.
- Develop animated agents as interesting and believable individuals with appropriate physical and psychological states, traits, and transitions, who express their unique identity in every aspect of their beings and behavior.
- Create a rich inventory of behaviors for the agent to choose from and give the agent the ability to perform the same behavior in different ways, modulated by its internal or external situation.
- Make sure that the agent does its best to avoid behavior that could offend users of a particular cultural group. Have the agent observe the sociocultural conventions determining when and how to interact with the user.
- Provide multiple, preferably bidirectional channels for the interaction between user and agent. Try to take advantage of verbal and non-verbal communication channels, whose use comes naturally to people.
- Equip the agent with explicit behaviors, protocols, and strategies for building, maintaining, and changing the relationship with the user.

Schaumburg's study of user responses to the Office Assistant (Schaumburg 2001), whose results were reported in the previous section, has two major implications for agent design. First, agent interfaces should be designed with a focus on assistance with the user's task first and social mimicry second. Social traits that prevent users from achieving their goals will likely be seen as a source of distraction and induce users to reject an agent. The user's ability to understand, predict, and control the agent's actions is critical (see also the comments on "role awareness" in the previous section).

Second, interfaces that seem human-like but do not adhere to the rules of social conduct are annoying to users. In the study, respondents expressed strong negative feelings toward the Office Assistant because it displayed behaviors that people regard as irritating or impolite in interactions with other humans (see "cultural awareness" above). Therefore, Schaumburg recommended that interfaces designed to display social behavior should be subjected to thorough long-term field testing with users to probe users' acceptance of these interfaces over time.

8.4 Summary and Outlook

There are countless ingredients that make up the human body and mind, like all the components that make up me as an individual with my own personality. Sure, I have a face and voice to distinguish myself from others, but my thoughts and memories are unique only to me, and I carry a sense of my own destiny. Each of those things are [sic!] just a small part of it. I collect information to use in my own way. All of that blends to create a mixture that forms me and gives rise to my conscience. (Major Motoko Kusanagi, [INT 59]).

The previous sections applied the design principles established in Chapter 7 to identify reasons for users' rejection of two earlier, less than successful embodied interface agents, Microsoft Bob and the Office Assistant (also known as the Paperclip). For each of these two agents, a detailed analysis was provided in Section 8.1 and Section 8.2, respectively, that showed which of the principles were violated in their design, and how. Based on the insights from this analysis, a number of lessons for agent design were compiled in Section 8.3. This final section builds a bridge to the following chapters by identifying the various ingredients (components) that have to be designed for embodied interactive software agents.

Any individual playing a specific social role in a certain group has many facets which define him or her as a person, as an occupant of the role, and as a party in various relationships with other members of the group, including such diverse aspects as hair style, voice, expertise, and interpersonal skills. All these aspects make an impact of their own but also contribute to the overall picture formed by others of the individual's identity, capabilities, and relationship to others. Likewise, embodied interactive agents are also social actors which are embedded in computer-based environments to serve particular functions or play specific roles in the user's experience, for example roles related to education (cf. Chapter 16). Diverse components have to be designed and combined in order to create an agent which, in the user's perception, is competent in its designated role, believable as an individual, and oriented toward building a relationship with the user. For pedagogical agents, this means that the finished agent should achieve three goals:

- It effectively *functions* in its appointed instructional role.
- It believably creates and maintains an attractive agent *persona*.
- It successfully establishes a productive, cooperative *relationship* with the user.

While traditional intelligent software agents are concerned with function, characters in films and games emphasize persona and virtual pets (cf. Chapter 7.16) focus on the relationship with the user. However, either function or persona or relationship alone provides an incomplete model for people and for embodied pedagogical agents as well. A complete model (cf. Chapter 7.6) combines all three aspects but carefully balances them so as not to make the agent seem unprofessional, boring, or aloof. "[A]nimate characters must complement function with persona – distinctive qualities of individual identity, personality, and psychology" (Hayes-Roth & Doyle 1998: 204), as well as with relationship-building behaviors.

Similarly, Gulz and Haake noted that there is a connection between the function and the visual appearance of animated pedagogical agents to the effect that agents with a poor visualization fail to engage and motivate learners because of their lack of appeal (Gulz & Haake 2006: 335). In the same paper (p. 325), they listed the following design elements of animated pedagogical agents. One element, expertise, was not discussed and has been added to complete the list. For each component, the corresponding chapter of the thesis is provided:

- Visual form and look (cf. Chapter 9);
- Movement characteristics (cf. Chapter 10);
- Facial expressions (cf. Chapter 10);
- Voice characteristics (cf. Chapter 10);
- Dialogue and conversational characteristics (cf. Chapter 11);
- Expertise (cf. Chapter 12);
- Emotion modeling, interpretation, and expression (cf. Chapter 13);
- Personality (cf. Chapter 13).

Two further chapters will discuss aspects of the implementation and evaluation of embodied interactive software agents. Chapter 14 will be concerned with agent architectures and technical platforms. The evaluation of embodied interactive agents will be the topic of Chapter 15.

9 Embodiment

For computational systems that are embedded in an environment and interact with people, the following issues become important (Cassell 2001: 67):

- How does the system represent itself in its user interface?
- How does the interface convey information about the environment and its own internal structure and operations to its users?
- How can users interact with the system through its interface?

The vision of the ‘disappearing’ or ‘invisible’ computer ([INT 60]; Norman 1998; Cai & Abascal 2006) involves doing away with a representation of the system by unobtrusively merging computer technology with objects and locations in the real world, up to a degree where no point of interaction is visible anymore. However, human beings are equipped for embodied interaction, and they rely on it in dealing with a complex and dynamic world (Cassell 2001: 68). Hence, they expect to be able to first locate intelligence and then interact *naturally* with it, i.e. without requiring special training or abilities and only relying on the means, experience, and knowledge they have regarding interactions with intelligent entities in the real world.[85] The best example of locatable intelligence that invites *natural interaction* (Valli 2004) is the (human) body (Cassell 2001: 68). By analogy, embodied interactive agents can help users to locate intelligence in an interface by providing an easily identifiable place for information and assistance. Being portrayed as animated characters, these agents can be perceived as individuals and interact with users as such, without evoking the impression of interacting with a machine. Articulate virtual bodies enable agents to mimic conversational behaviors from human face-to-face interaction, including speech, gesture, gaze, posture, intonation, and other verbal and non-verbal behaviors (cf. Chapter 10). A well-designed embodiment also communicates an agent’s identity, giving users a good idea of what sort of person the agent portrays, where it comes from, how it might behave, etc. Embodied agents with appealing and interesting attributes and features can attract users and make it easier for them to suspend their disbelief. Finally, agent embodiments portraying a believable and likeable persona can aid considerably in building user affection and trust, which are important for developing long-term friendly relationships with users (cf. Chapter 7.16). People have been found to perceive embodied agents as more friendly, helpful, and polite (McBreen & Jack 2001), and they seem to enjoy and benefit from the experience of learning with them (Lester et al. 1997a; Lester et al. 1997b; Lester et al. 1997c; Moreno et al. 2000; Atkinson 2002). Furthermore, natural interaction in general occurs at a lower level of abstraction than more artificial forms of interaction. Since higher levels of abstraction require more cognitive effort for the interaction as such (Valli 2004: 8), natural interaction can help to reduce the user’s cognitive load (cf. Chapter 2.2.2) and thus free resources for attention on content (p. 11).

The digital representation of an embodied interactive agent in its environment is the first characteristic that is perceived and interpreted by users. The agent’s embodiment exhibits cues

---

[85] Natural interaction relies on perceptions, processing, and actions that (a) humans are “made for,” i.e. which are “implicitly written in the structures of minds and bodies” (Valli 2004: 8) (e.g. use of the hands to manipulate objects); or (b) are of cultural origin but have become so engraved in the minds of people that they are viewed as a part of the human heritage (e.g. deictic gestures, cf. Chapter 10.1.2). Both types have in common that people apply them all the time in their real lives, without thinking and without the need for explicit instruction.

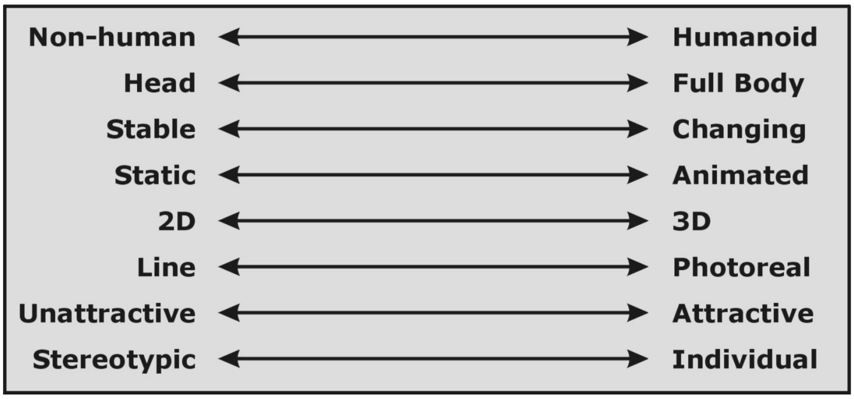

Figure 35. Dimensions of agent embodiment. Adapted from Churchill et al. (2000: 76, Figure 3.3).

that indicate the agent's availability for interaction, as well as its identity, traits, domain, and capabilities. Making the right first impression on the user should thus be a central concern for the designers of embodied interactive software agents. Cues conveyed by the agent's appearance and behavior should give users a clear idea of its qualities and limitations (Churchill et al. 2000: 74). Designing an appropriate embodiment for a particular agent involves making choices from a multi-dimensional space of design options (cf. Figure 35). The various dimensions of agent embodiment are discussed in the following sections.

9.1 Personification

"The outward appearance is dependent on a robot's function. My own function requires a very manlike appearance, and I have it. Others are different, although all are humanoid. Certainly they are more humanoid than the distressingly primitive models I saw at the shoe counter. Are all your robots like that?"
"More or less," said Baley. "You don't approve?"
"Of course not. It is difficult to accept a gross parody of the human form as an intellectual equal. Can your factories do no better?"
"I'm sure they can, Daneel. I think we just prefer to know when we're dealing with a robot and when we're not." (Asimov 1997: 54).

"Why the human form?"
"Because the human form is the most generalized form in all nature. We are not a specialized animal, Mr. Baley, except for our nervous systems and a few odd items. If you want a design capable of doing a great many widely various things, all fairly well, you could do no better than to imitate the human form. Besides that, our entire technology is based on the human form. (...) It is easier to have robots imitate the human shape than to redesign radically the very philosophy of our tools." (Asimov 1997: 134).

When designing the body for an animated agent, an appropriate representation has to be chosen: should the body represent a person, some other living being, or an inanimate object (Ruttkay et al. 2002)? Agent embodiments may be more or less similar to humans in appearance (Churchill et al. 2000: 75). At one end of the continuum, the agent is portrayed as a *virtual human* (Badler et al. 1993; Badler et al. 2000; Gratch et al. 2002; Swartout et al. 2006) featuring the central elements of a humanoid embodiment (Churchill et al. 2000: 75ff.):

- *Face*. Facial expressions convey cognitive and affective states (cf. Chapter 10.1.3).
- *Hands*. The hands perform conversational and propositional gestures (cf. Chapter 10.1.2).
- *Feet*. The feet provide the capability of movement (cf. Chapter 10.1.6) and indicate the direction of the agent's attention.
- *Torso*. Body posture can signal involvement and orientation (cf. Chapter 10.1.5).

However, apart from complete figures (cf. Figure 36a), humanoid embodiments can also be partial bodies lacking feet or a torso (cf. Figure 36b).[86] At the other end of the continuum, the agent may present itself as an alien creature, an inanimate object, or an abstract shape (cf. Figure 36c–e) (Churchill et al. 2000: 75). Between these extremes, agent embodiments may integrate a variety of human features into a non-human representation.

Designers may choose to portray embodied agents in the image of humans in order to exploit *anthropomorphism* (or *personification*), the general tendency of people to attribute human-like characteristics (such as emotions and sophisticated planning capabilities) to non-human organisms or objects. The anthropomorphic perceptions and ideas of people influence how they interact with animals, inanimate objects, and possibly embodied agents.

Design in general has a long history of imitating the human form in all kinds of artifacts ([INT 61]), for reasons of usability (cf. Chapter 6.3), attractiveness (cf. Chapter 6.4), and social ability (cf. Chapter 6.5). In the design of current everyday products, the anthropomorphic form has four distinct uses (DiSalvo & Gemperle 2003):

- *Make things familiar*. The anthropomorphic form makes the purpose and functioning of a new product transparent and reduces users' reluctance to use or purchase it.
- *Keep things the same*. The anthropomorphic form may have become a defining characteristic for a certain class of products, so changing the form may lead to confusion about the identity, function, or purpose of the product.
- *Reflect product attributes*. The anthropomorphic form indicates the features and limitations of the product, thus guiding people in how they can interact with and relate to it.
- *Project human values*. The anthropomorphic form is a way to express personal, social, or cultural values related to the product or its applications.

Whether anthropomorphic forms should be used in the design of user interfaces is the topic of an ongoing debate (Laurel 1990b; Don et al. 1992; Laurel 1992; Erickson 1997; Morgan 1995; Shneiderman 1995; Shneiderman 1997; Shneiderman & Maes 1997; Duffy 2003; Zimmerman et al. 2005). Proponents of anthropomorphism emphasize that it facilitates interaction because people automatically apply their knowledge about face-to-face interactions and social relationships with other humans.[87] An expressive and sensitive anthropomorphic

---

[86] In contrast, the face and hands are critical elements of embodied interactive agents participating in face-to-face conversations with the user. Facial displays and gestures help to disambiguate messages, structure the conversation, and add information that is not communicated by speech (Cassell 2000a: 6). Furthermore, both the face and the hands are among the principal carriers of affective information (Lester et al. 2000: 141).

[87] It should be noted, though, that Nass and Moon (2000) rejected anthropomorphism as an explanation for human social responses to computers. Instead, they argued that people *mindlessly* (Langer 1989; Langer 1992) rather than reasonably apply social rules and expectations to computers. The subjects in their experiments did not believe that computers were essentially human and vehemently denied that they had responded socially to them during the experiments, which, in fact, they had (Nass & Moon 2000: 93).

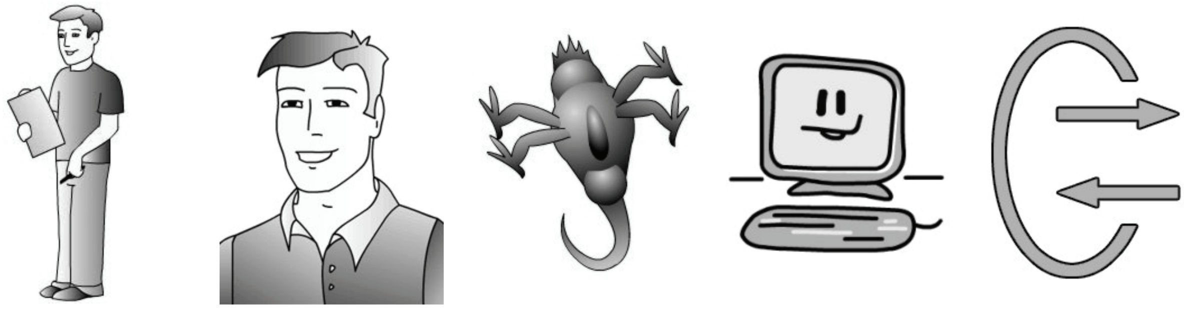

Figure 36a–e. Examples of different personifications, including (from left to right) a full human body, a partial human body (head and upper torso), an alien creature, an inanimate object (with human features), and an abstract shape (all images taken from the Virtual Linguistics Campus ([INT 3])).

agent is claimed to help users to feel more comfortable with, communicate more effectively with, and relate more readily to the agent. Those opposed to the idea have argued that anthropomorphism can evoke incorrect expectations of human competencies and establish dangerous dependency relationships with machines (cf. Chapter 3.2.5).

Carefully weighing the pros and cons of using the anthropomorphic form is a crucial consideration in the design of effective human-computer interfaces. The following positive aspects of anthropomorphic interfaces should be mentioned:

- Anthropomorphism can elicit social and emotional responses in users, which are instrumental in forming a relationship between them and the machine. For example, Brennan and Ohaeri (1994) found that users of a (simulated) natural language (NL) interface treated the computer more as a social partner when the messages of the NL interface were 'anthropomorphic,' consisting of complete grammatical sentences that used first person pronouns whenever possible (cf. the personalization principle in Table 1). Anthropomorphic messages also led people to use more indirect requests and conventional politeness in their responses. However, they did not attribute greater intelligence to the anthropomorphic interface. In contrast, the participants in the study conducted by King and Ohya (1995) rated anthropomorphic forms (faces) as possessing a higher degree of agency and intelligence. The results of Parise et al. (1999) showed that subjects were more likely to cooperate with a computer character if it resembled a person, whereas they tended not to cooperate with and keep their promises to a non-human character. But a more likeable character did not lead subjects to cooperate with it. Parise et al concluded that "static assessments of character attributes (such as attractiveness, likeability, or trustworthiness) are inappropriate and can be misleading when people encounter those characters in a social interaction framework" (Parise et al. 1999: 140).
- Anthropomorphism exploits a natural and automatic tendency in humans. People are adept at relating to and communicating with other people, and they tend to see human attributes in a shape or action appearing in the least bit human-like or intelligent (Heider & Simmel 1944).
- Anthropomorphism is the application of a metaphor (Laurel 1990b).[88] Metaphors provide a correspondence or analogy with real-world objects which are familiar to the user, and thus

[88] It has been argued that anthropomorphic metaphors are very common in the domain of computing (Johnson 1994). People commonly relate well-known activities, processes, and states associated with human beings to the computer in order to reduce the complexity which they perceive in machines to a manageable level (Marakas et al. 2000: 722). The vocabulary used to describe the capabilities,

help him or her to build a mental model of a system (cf. Chapter 6.3). Anthropomorphism is an important element of designing interfaces that invite conversational interaction and allow people to apply their existing models and skills of face-to-face conversational interaction. Anthropomorphic interfaces exploiting real-world analogies and metaphors can help to remove the barrier between the user and the machine.
- Social interaction may be more effective than direct manipulation in tasks that involve delegating complex or redundant actions or accomplishing things beyond the current place and time (Don 1992). Conversational interfaces thus have their place in human-computer interaction, whether they portray a human-like figure or not. In any case, they should behave as coherent (cf. Chapter 7.4) and consistent (cf. Chapter 7.5) interaction partners, which make their limitations (cf. Chapter 7.14) transparent.

However, there are also valid arguments against using the anthropomorphic form in human-computer interfaces. Some of them are listed below:

- Anthropomorphic interfaces fail to maintain the user illusion because they are unable to support their projected intelligence with real capabilities, which, in turn, leads to user disappointment (Kay 1984). Anthropomorphism thus creates a dilemma for human-computer interaction: it is meant to improve the interaction with a computer, but the user should not attribute intelligence to the machine which it does not (yet) possess. This, however, is a natural and automatic response. Current computers are 'semi-intelligent' at best, but it is not clear how this can be adequately portrayed (Erickson 1997).
- Human-computer interaction is a partnership in which both users and computers act on and respond to each other. However, true social interaction (cf. Chapter 5.4.5) presupposes that all parties involved possess sophisticated knowledge and inference capabilities (cf. Chapter 12) as well as communicative competence (cf. Chapter 10.1 and Chapter 11), all of which are still underdeveloped in computers (Erickson 1997). Interfaces suggesting that they could hold up their end of a social interaction but fail to do so cause disappointment and frustration.
- Anthropomorphic agents may become so 'realistic' (human-like) that users cease to perceive them as useful. Instead of trusting in the speed, accuracy, and reliability of the machine, they might come to expect it to be slow, inaccurate, and unreliable, just like real people (Foner 1993; Morgan 1995).
- Not all users may be comfortable with human-like machines (cf. the first quotation at the beginning of this section). Hence, an anthropomorphic interface may appeal to some users but create anxiety or confusion for others.
- People who repeatedly get away with abusing agent assistants, librarians, tutors, etc. that are human-like in appearance and behavior might be encouraged to treat real people in the same roles in the same way (Laurel 1990b).[89]

---

operation, and states of computers is full of anthropomorphisms. For example, computers 'read,' 'think,' and 'cheat,' and they may be 'friendly,' 'intelligent,' or "infected with a virus." Marakas et al. argued that in constructing this vocabulary, people fall back to what is most familiar to them, namely themselves (p. 722).

[89] Bartneck et al. (2005) investigated the question if humans abused a robot to the same extent as they abused another human being, by replicating a series of experiments on obedience originally performed by Milgram (1974). Subjects were asked to teach a robot student word combinations and to punish mistakes made by the robot with an electric shock. As the voltage increased, the robot

While the critics of anthropomorphism may be right about these issues, it still seems that the advantages of using the human form in the right way outweigh the potential pitfalls:

- Anthropomorphism is a ubiquitous, inevitable tendency of people, so if one cannot beat it, why not join it and make it useful?
- People have been shown to respond to ordinary computers as if they were social entities, even when the machine presented itself only through text messages (cf. Chapter 6.5). From this point of view, adding a face and a body to a user interface does not seem revolutionary but more like the next logical step.[90]
- Designers can effectively make use of anthropomorphism if they understand its strengths and weaknesses (Don 1992). On the one hand, it is a fact that at this point, little is known about people's reactions to anthropomorphic interfaces, but if such interfaces are not built, deployed, evaluated, and improved in the context of realistic scenarios, their strong and weak points will never be understood.
- Using the anthropomorphic form for the sake of cuteness and entertainment alone is unlikely to work in an interface. To actually improve human-computer interaction, anthropomorphic interfaces have to show behavior that is functional with regard to the system's aim (Dehn & van Mulken 2000: 20).
- The true potential of anthropomorphic interfaces lies in the possibility of rendering the interaction multimodal. An animated on-screen character can be used to portray certain communicative processes, including averting gaze, establishing eye contact, gesturing, and posture. Such embodied conversational agents (cf. Chapter 3.3 and Chapter 11.3.4) have the same properties as humans in face-to-face interactions (Cassell 2000b: 72):
  - They can interpret the verbal and non-verbal behaviors of others.
  - They can respond with their own verbal and non-verbal behaviors.
  - They can regulate the conversation through turn-taking, feedback, repair mechanisms, etc.

---

would react with displays of increasing pain and eventually beg that the shocks be stopped, communicating through facial expressions of emotion and messages delivered using synthetic speech. Although the participants showed compassion for the robot, urged on by the experimenter, all of them eventually applied the strongest (deadly) electric shock, whereas only 40% did the same in Milgram's original study involving a human actor portraying the effects of the simulated electric shocks. Bartneck et al. concluded that people may have fewer concerns abusing robots than they have abusing people. Slater et al. (2006) also repeated Milgram's experiment, but they used a female virtual human to play the role of the learner. They found that although all participants knew that neither the learner nor the electric shocks were real, a significant number of those who saw and heard the virtual human during the experiment showed care for its well-being and felt bad about administering the shocks. Slater et al. concluded that humans interacting with virtual characters in an extreme social situation tend to respond to the situation at the subjective, behavioral, and physiological levels as if it was real. However, most participants who saw and heard the virtual human still applied all or almost all of the shocks, despite their misgivings, which confirms the results obtained by Bartneck et al. that the threshold for abuse seems to be lower for virtual humans than for real humans.

[90] On the other hand, it could be argued that since minimal cues, such as text, are sufficient to induce a social response in users, there seems to be no need at all to include (anthropomorphic) embodied characters in human-computer interfaces (Dehn & van Mulken 2000: 3). However, character-based interfaces can be beneficial to human-computer interaction in various ways (cf. Chapter 3.1.4).

- They can indicate the current state of the exchange and contribute new information by means of audio-visual cues.

- Adequately designed embodied anthropomorphic interfaces can benefit human interaction with computers in ways that go beyond the capabilities of current keyboard- and mouse-based user interfaces (Cassell 2000b: 72):
  - They increase the robustness of human-computer dialogues despite the shortcomings of speech recognition (cf. Chapter 11.1.1) and natural language understanding (cf. Chapter 11.1.3) technologies.
  - They add further channels for human-computer interaction, providing a low-cost way to increase communication bandwidth between human and machine.
  - They facilitate both human-computer interaction and (avatar-based) interaction between humans mediated by technology.

The decision to include specific human features should depend on the overall embodiment chosen, the range of behaviors[91] intended for the agent, and how much the agent should appear human-like or individual. Zimmerman et al. (2005) suggested that the design choice between a human, animal, or iconic form as the embodiment of a specific agent should be influenced by the proximity of the agent's role to real-world social roles played by people, as it is perceived by the user. If users lack a clear mental model of a human in the role portrayed by an agent, it is more feasible to use a non-human embodiment. This may not be an option for pedagogical agents, though. Since most learners have a fairly clear mental model of a human instructor from their previous experiences, they may not find a non-human embodiment credible for an agent in an instructional role.

In general, designers of agent-based interfaces must be careful not to create representations for agents that are inconsistent with their identity, role, and capabilities (cf. Chapter 7.5). Agents should not be personified in such a way that users are misled,[92] which could mean choosing an obviously non-realistic representation for the agent or even none at all (Wexelblat & Maes 1997, cf. Section 9.5). Furthermore, they need to be aware of any stereotypes which a particular representation might evoke in the target audience. If a human representation is chosen, it could be a well-known real person, represent a category of real persons (defined in terms of age, gender, ethnicity, etc.), or a new individual could be created (Ruttkay et al. 2002). Each of these options invariably evokes a set of stereotypes associated with the person or group portrayed (Haake & Gulz 2008: 2), and users will expect agent behavior (cf. Chapter 10) to be consistent with those stereotypes. On the other hand, integrating a few traits that go *against* common stereotypes can be useful in order to build characters that are both intriguing and memorable (Isbister 2006: 14).

---

[91] For example, a mouth is necessary to speak, hands are needed to grasp objects, etc.

[92] Nowak and Biocca (2003) examined how users' sense of presence in a virtual environment was influenced by the perceived agency of interaction partners (human-controlled avatar or computer-controlled agent) and the level of anthropomorphism of the image used to represent them. While participants responded socially to both human- and computer-controlled entities, the findings of the study also indicated that "the more anthropomorphic images set up higher expectations that [led] to reduced presence when these expectations were not met" (Nowak & Biocca 2003: 481, cf. Nowak 2004). Nowak and Biocca recommended that "increasing anthropomorphism (…) should be done only when the interface and system can meet higher expectations" (p. 492). This recommendation is in line with the principle of balanced design discussed in Chapter 7.3.

9.2 Appearance

Appearance is the one aspect of agent embodiments that attracts the user's attention first. It influences how effectively and credibly an agent can fulfill its role and guides users in their interactions with the agent (Maldonado & Hayes-Roth 2004: 155). Toby Gard, designer and lead artist of the popular computer game *Tomb Raider* ([INT 62]) and its main character, *Lara Croft*, argued that

> *a person's first impression of a character will almost certainly come not from what they do, think, or say, but what they look like. If the character makes a good first visual impression, players will likely stay focused on it, allowing you to further entice them with the character's personality. (Gard 2000: 4f.).*

People tend to form beliefs (cf. Chapter 12.1) about their human interaction partners, especially their first impression of a stranger, by interpreting their observable appearance and behavior based on stereotypes (cf. Chapter 7.15).[93] These beliefs allow them to anticipate the behavior of others, which in turn drives their own behavior (Churchill et al. 2000: 74). As Haake and Gulz put it, "visual stereotypes frame our expectations" (Haake & Gulz 2008: 4). They serve as cognitive aids that help people to navigate complex social environments (p. 4).

The intuition of folk psychology that "first impressions are lasting impressions" is supported by empirical research which found that attraction based on appearance (and personality information) both persists and deepens over time (Mathes 1975; Green et al. 2008: 2457). Given that people treat computers (and embodied agents) like they treat humans (cf. Chapter 6.5), the principle of lasting first impressions may also apply to embodied agents (Gulz & Haake 2006: 328). Indeed, Buisine et al. (2004) found a main effect of the appearance of an embodied presentation agent on the agent's likeability. This effect was independent of the participants' gender. The results also showed a tendency of appearance to affect recall performance, but that effect did not reach statistical significance. The findings of Parise et al. provided "strong evidence that the mere appearance of a computer character is sufficient to change its social influence" (Parise et al. 1999: 136).

Embodied interactive agents should therefore be designed to engage users already through their appearance, which includes various aspects of an agent's physical form that are listed in Table 12.[94] Some guidelines for designing appealing agents include:

- Help users to immediately understand the agent's *identity*, *role*, and *actions* (cf. Chapter 7.7).
- Make the agent as *physically attractive* as possible for the target audience (Maldonado & Hayes-Roth 2004: 155). Consider hair style, cosmetics, and costume as important visual clues (Gulz & Haake 2006: 332). But keep in mind that "[e]valuations [of attractiveness]

---

[93] Fogg made the same point for computing products, stressing the importance of surface traits for helping users to form an initial judgment (of a product's credibility) based on a quick first-hand inspection of these traits (Fogg 2003: 132f.).

[94] Buisine et al. (2004: 233) found anecdotal evidence for the effect of clothing and accessories on likeability. One of the agents in their experiment was disliked by some of the participants for wearing a white coat which made the agent appear "too strict," although others thought of the same agent that it was nicer and more serious. Another agent wearing glasses elicited negative responses because users could not clearly see the agent's eyes through the glasses.

Table 12. Aspects of agents' appearance.

Species	**Demographics**
• *Human* • *Animal* (e.g. bear, dog, parrot, etc.) • *Fictitious* (e.g. alien, fictional or mythical creature) • *Object* (e.g. car, computer, paperclip, plane, pencil, etc.) • *Shape* (e.g. square, circle, triangle, etc.)	• *Age* (same or different age group as the user) • *Sex* (male, female, or none) • *Ethnicity* (similar to or different from the user's ethnicity)
Body Features	**Outfit**
• *Build* (e.g. tall .. short, fat .. thin, strong .. weak, etc.) • *Eyes* (size, shape, and color) • *Hair* (style and color of facial and scalp hair) • *Skin* (pale, mid-toned or dark complexion) • *Peculiarities* (e.g. scars, tattoos, deformed body parts, etc.)	• *Clothing* (e.g. none, formal .. casual, traditional .. modern, professional .. leisure, etc.) • *Accessories* (e.g. bags, glasses, jewelry, tools, weapons, etc.)

vary from culture to culture, generation to generation, and individual to individual" (Fogg 2003: 94).[95]

- Design the agent with a *human-like face* as a means of engaging the user. *Eyes*, *mouth*, and *eyebrows* are essential for conveying the agent's cognitive and emotional state (Ekman 1982; Wilkinson 2006). Given that state-of-the-art computer graphics and processors are powerful enough to render increasingly visible and subtle facial expressions, the design of the faces of animated characters requires more attention (Isbister 2006: 143, cf. Chapter 10.1.3).
- Give the agent an *ethnicity* that users perceive as *similar* to their own to increase its perceived attractiveness, trustworthiness, and persuasiveness (Nass et al. 2000a).
- Mind the impact of *gender roles*, *cultural norms*, and *traditional attires*. Role assignment is subject to gender stereotypes. Accepted behaviors differ for men, women, and children. There are also dress codes for different roles, groups, and occasions (Maldonado & Hayes-Roth 2004: 155).
- Prefer *neutral* representations for cross-cultural agents that do not inspire strong negative responses in any culture over representations that are strongly preferred in some cultures but strongly rejected in others. But note that agents that are too bland and adaptive have no appeal anywhere (Maldonado & Hayes-Roth 2004: 157).
- Find a *balance* in design (cf. Chapter 7.3) where members of the represented culture can recognize themselves in an agent persona and, at the same time, members of other cultures can correctly identify the cultural background portrayed by the agent, without offending either group (Maldonado & Hayes-Roth 2004: 152).

Zimmerman et al. (2005) found that people's preference for male vs. female agent forms is influenced by the gender stereotypes which they have for particular tasks and social roles. For example, the participants in their study were affected in their judgment by the gender stereotype

[95] Still, it seems that some aspects of attractiveness, such as health, symmetry of face and body, and a straight profile, are predictable (cf. Chapter 6.4).

of women playing supporting roles and thus rated female agents higher for the roles of receptionist, realtor, and librarian. Moreno et al. (2002) reported that learners applied gender stereotypes to animated anthropomorphic tutor agents and that the expectations associated with these stereotypes affected their learning. The tendency of people to transfer stereotypical expectations with respect to females and males in human relationships to their interactions with pedagogical agents is also apparent from results obtained by Baylor and Kim, according to which both female and male college students were more motivated by male instructor agents than by female instructor agents, perceived the male agents more favorably (Baylor & Kim 2003b; Baylor & Kim 2004; Kim & Baylor 2005a), and showed higher recall performance after working with a male agent (Kim & Baylor 2005b). Likewise, the research by Baylor and Plant (2005) showed that consistent with gender stereotypes and the *similarity attraction principle* (people are more attracted to other people who they see as similar to themselves),[96] female undergraduate students identified most with agents that were female, attractive, young, and 'cool;' thought that the prototypical engineer was represented by a male, old, and 'uncool' agent; and preferred to learn about engineering from male agents that were attractive but uncool. However, further research by Baylor et al. (2006) and Rosenberg-Kima et al. (2008) indicated that agent appearance can also be used to change stereotypical beliefs. In the study by Baylor et al., an agent with the same gender as the participants (female) and representing itself as an engineer (commonly viewed as a less stereotypical role for females) was more influential than a male agent for inducing a more positive stereotype of engineering in the young female subjects. Still, the male agents in the study were more effective than their female counterparts in motivating young women to pursue an engineering career and in changing their views of the usefulness of engineering, possibly because they were perceived as the prototypical engineers and as the individuals with greater expertise.[97] Baylor et al. concluded that in applications designed to change attitudes and beliefs (i.e. to persuade, cf. Fogg 2003), it is important to employ both agents that are perceived as peer models and agents that are perceived as experts.

Concerning the impact of the ethnicity portrayed by an agent, Baylor and Kim (2003b) reported that college students preferred to interact with agents having the same ethnicity as their own. This effect was most apparent with African American learners, who showed a significant tendency to choose an agent instructor with an African American ethnicity and to perceive the chosen agent instructor more favorably (Baylor et al. 2003). Earlier research by Nass et al. (Lee & Nass 1998; Nass et al. 2000a) indicated the same basic preference: an embodied agent of the same ethnicity as the user was rated higher in terms of similarity to the user, social attractiveness, and trustworthiness than one with a different ethnicity (Nass et al. 2000a: 382). An explanation for these results may be that place of origin (ethnicity), like gender and personality, can be a strong predictor of the attitudes and behaviors that are likely to be displayed by a conversational partner (Nass & Brave 2005: 61). However, it appears that results with respect to the ethnicity of agents are not consistent, as Moreno et al. (2002), for example, found no effects of the ethnicity of pedagogical tutor agents on learners' stereotypical expectations or learning. In a later paper, Moreno and Flowerday (2006) reported that from among the participants in their study, African American, Asian American, Hispanic, and Native

[96] On similarity attraction in human-computer interaction, see Reeves and Nass (1998: 89–99), Nass and Brave (2005: 34–38), and the references cited by these authors.

[97] In general, there are both stereotypically 'female' (e.g. cooking, sewing, and child care) and stereotypically 'male' subjects (e.g. hunting, car repairs, and computer programming), and learners tend to think that both female and male teachers are better when they operate within their stereotypical domains of expertise (Nass & Brave 2005: 29).

American students were significantly more likely to choose a pedagogical presentation agent (cf. Chapter 12.3.12) of the same ethnicity as their own than White American students. Interestingly, participants who chose a *different*-ethnicity pedagogical agent outperformed those who learned from a same-ethnicity agent because the latter focused more on how the pedagogical agent represented *them* than on the presentation (Moreno & Flowerday 2006: 204).

9.3 Individuality

Embodied agents should seem individual (cf. Chapter 7.8), i.e. they should have their own unique and distinctive persona, which is revealed and expressed in every aspect of their beings and behavior, including their embodiment. Individuality manifests itself in the appearance of an agent (cf. Chapter 9.2), on the one hand in relatively stable physiological components (facial features, build, deformities, etc.) and on the other hand in variable additions and modifications that reflect the style and taste of the individual, including accessories, clothing, hair style and color, and makeup. The latter aspects of the agent's embodiment can (and should) vary in the course of a series of encounters between the agent and users (cf. Chapter 16.3.1).

An individual embodiment cannot be designed without knowing the identity of the agent it will portray (Thomas & Johnston 1984: 247). So the first task of agent designers, similar to the one of writers, is to create a profile of the character which they intend to develop. The *character profile* describes various aspects of the character, including the following ([INT 63]):

- *Demographic*. Name, age, sex, nationality, residence, occupation, income, etc.;
- *Physical*. Build, eye/hair/skin color, disabilities, health, etc.;
- *Intellectual*. Educational background and intelligence level;
- *Mental*. Mental strength, mental stability, mental illnesses, etc.;
- *Talents/skills*. Artistic, manual, social, technical, etc.;
- *Personality*. Introvert or extrovert, qualities and flaws, self-perception and self-confidence, short-term and long-term goals; etc.;
- *Habitual*. Dependencies (smoking, drinking, etc.); idiosyncratic speech patterns, gestures, and other behaviors;
- *Emotional*. Causes of anger, fear, happiness, etc. and ways of handling these emotions;
- *Spiritual*. Nature and role of the character's spiritual beliefs;
- *Style*. Elegant .. shabby, casual .. formal, conventional .. unconventional, etc.;
- *Experience*. Formative events and influences in the character's life.

Rowe et al. (2008) suggested the use of *character archetypes* as blueprints for character profiles. A character archetype defines a set of traits (including fears, goals, motivations, and personality characteristics, among others) that are associated with a given character and role. Rowe et al. argued that archetypes can be used to create characters with consistent sets of traits that are both familiar and believable to users.

One of the most important sources of inspiration for character designers is the real world with its wide diversity of people. Designers should train themselves to constantly watch out for the unusual, interesting, and entertaining in the people around them (Thomas & Johnston 1984: 247). However, they also need to pay attention to the requirements imposed by the targeted role (e.g. advisor, guide, opponent, teacher, or salesperson), environment (e.g. corporate identity, technical platform, etc.), and audience (gender, age, culture, education, social class, occupation, organizations, limitations, skills, goals, needs, expectations, experiences, and preferences, cf. Chapter 6.1).

Embodied agents can be designed to portray different degrees of individuality. Since people commonly draw on stereotypes about groups of people to form opinions and expectations about others, agent embodiments can purposefully evoke these stereotypes to guide users' behavior. For example, a motherly figure gives rise to expectations of a warm and nurturing caretaker while a persona with 'typical' Asiatic features would appear to be from an Asian culture. Here, stereotypes are positive constructs that describe patterns of diversity in human characteristics which facilitate interaction, with no intention to offend the represented culture (Maldonado & Hayes-Roth 2004: 152).

9.4 Movement

A central consideration for the design of embodied agents concerns the extent to which a chosen embodiment can move (Churchill et al. 2000: 77) in order to act on and communicate with other entities. The continuum ranges from static images that do not move at all to fully animated figures capable of expression, gesture, and locomotion. While non-living objects can use movement to communicate messages, many current embodied agents possess an animated humanoid embodiment (cf. Figure 17a to Figure 17d) to take advantage of the fact that the human body and embodiments with similar configuration and capabilities are well equipped for multimodal embodied communication and thus leverage users' natural skills in interpreting and generating body movements (Churchill et al. 2000: 77; Cassell 2000a; Cassell 2001, cf. Chapter 3.1.4).

Movement of embodied agents is implemented through *animation*, the art of creating the illusion of motion by displaying a series of slightly shifted images ('frames') in rapid succession. While in traditional animation the frames are hand-drawn, today, animations are commonly created and displayed by using computers.[98] Character animation (Thomas & Johnston 1984; Maestri 1996; Maestri 1999; Maestri 2001) is the subfield of animation that builds the behaviors of artificial characters to make them come alive for the audience. Embodied agents may perform animations in interactions with users to support their particular role, enhance their own persona, and facilitate relationship building with the user (cf. Chapter 8.4). The first type of animations serves users in their current task, the second type serves the agent's believability (cf. Chapter 10.2), and the third type serves the user-agent relationship (cf. Chapter 10.3).

Baylor & Ryu (2003) tested the impact of static image and animation on how pedagogical agents were perceived in terms of various personal characteristics (person-like, engaging, credible, and instructor-like, cf. Chapter 4.2). From the results of their study, the authors recommended animation as (one of) the most effective ways to implement all investigated characteristics.[99] Neither positive nor negative effects of animation on performance were found, which may help to reassure those who fear that animated agents might distract the learner.

[98] Bétrancourt and Tversky defined *computer animation* as "any application which generates a series of frames, so that each frame appears as an alteration of the previous one, and where the sequence of frames is determined either by the designer or the user" (Bétrancourt & Tversky 2000: 313).

[99] However, Baylor and Ryu also stressed the importance of voice (cf. Chapter 10.1.1) for agent-based learning environments (on the benefits of the presence of voice, see also Reeves and Nass (1998), Moreno et al. (2001), and Atkinson (2002)). Later research by Rosenberg-Kima et al. (2007; 2008)

9.5 Realism

The options for the visual rendering of embodied agents differ with respect to their degree of realism and the 'depth' of the created images. The two extremes of the realism dimension are marked by photorealistic embodiments, which are indistinguishable from individuals portrayed in photographs or movies, and simple line drawings. Other ways to define this dimension include (Badler 1997):

- Realistic .. artistic representations;
- Physiologically accurate models .. cartoon shapes;
- Human limitations .. cartoon actions.

The field of computer animation is making rapid progress, as demonstrated by an increasing number of commercially successful movies that are completely computer-generated, such as *Toy Story* ([INT 64]), *Antz* ([INT 65]), and *Shrek* ([INT 66]), in addition to movies where *computer-generated imagery (CGI)* techniques are used to create animated characters playing opposite human actors (e.g. *Jurassic Park* ([INT 67]), *Star Wars Episodes I-III* ([INT 68]; [INT 69]; [INT 70]), and *The Lord of the Rings* ([INT 71]; [INT 72]; [INT 73])). Some of these characters, such as Gollum from *The Lord of the Rings*, have become very popular with their audience and have even won awards of their own while others, like Jar Jar Binks from *Star Wars Episodes I-III*, have stirred up considerable controversy and have even come to be widely rejected by moviegoers. All these movies have in common that their stories and/or characters are obviously fantastic, not from our world, or not from our time (Dautenhahn 2004: 51).

But what if computer-generated movies were created with the explicit goal to look as realistic as possible, featuring CGI humans whose very skin and hair look almost real? A recent example of a serious attempt at computer-generated photorealism, including photorealistic people, is the movie *Final Fantasy: The Spirits Within* ([INT 74]). Unfortunately, the movie disappointed audiences and critics alike because of the overuse of special effects, the one-dimensional characters, the emphasis of spectacle over plot, and, last but not least, the use of CGI to simulate live actors, which was criticized as an overly expensive gimmick. The virtual humans acting in the movie were advertized as photorealistic, but their movements and expressions were described as stiff and unexpressive. Particular shortcomings included a doll-eyed stare, rigid character postures and gaits, and the lack of visible perspiration and skin and tissue deformation during character movement. Finally, the characters looked like painted statues due to limitations of the lighting models used for skin and hair ([INT 75]). Overall, *Final Fantasy: The Spirits Within* nicely illustrates the tension between realism and believability in (character) animation: more realistic animations are not necessarily more believable (Dautenhahn 2004: 51).

Another illustration of the negative effects of a design aiming to be as realistically human-like as possible on people comes from the science fiction literature. In Isaac Asimov's classic robot novels *The Caves of Steel* (Asimov 1997), *The Naked Sun* (Asimov 1991), *The Robots of Dawn* (Asimov 1994), and *Robots and Empire* (Asimov 1996), one of the protagonists is R. Daneel Olivaw, where 'R' stands for 'Robot.' R. Daneel is a robot created in the human image, down to the smallest detail and with the highest level of realism that is technologically possible:

---

found evidence for the importance of the visual presence of an anthropomorphic interface agent in addition to voice.

[Robot] Daneel's skin texture was perfect, the individual hair on head and body had been lovingly and intricately manufactured and placed. The muscle movement under the skin was most realistic. No pains, however extravagant, had been spared. Yet [Plainclothesman Elijah] Baley knew, from personal knowledge, that limbs and chest could be split open along invisible seams so that repairs might be made. He knew there was metal and silicone under that realistic skin. He knew a positronic brain, most advanced but only positronic, nestled in the hollow of the skull. He knew that Daneel's "thoughts" were only short-lived positronic currents flowing along paths rigidly designed and foreordained by the manufacturer. (Asimov 1991: 27).

At first (and even second) glance, this robot passed for a human being, deceiving various human characters in *The Caves of Steel* (Asimov 1997: 25f, 79–89, 132–135). However, R. Daneel's human façade crumbled under closer scrutiny. Its literal-mindedness, unnatural manner of speech, posture and body movement, unemotional gravity, and even the very perfection of its human appearance[100] gave it away, albeit only to the more careful human observer:

Baley ate industriously but without the relaxation that allows complete enjoyment. Carefully, he flicked an occasional glance at R. Daneel. The robot ate with precise motions of his jaws. Too precise. It didn't look quite natural.
Strange! Now that Baley knew for a fact that R. Daneel was in truth a robot, all sorts of little items showed up clearly. For instance, there was no movement in an Adam's apple when R. Daneel swallowed. (Asimov 1997: 112).

"Humaniform robots are quite like human beings in appearance, Partner Elijah, down to the hairs and pores in our skin. Our voices are thoroughly natural, we can go through the motions of eating, and so on. And yet, in our behavior there are noticeable differences. There may be fewer such differences with time and with refinement of technique, but as yet they are many. You – and other Earthmen not used to humaniform robots – may not easily note these differences, but Aurorans would. No Auroran would mistake Jander – or me [R. Daneel] for a human being, not for a moment." (Asimov 1994: 39).

Those who eventually found out that R. Daneel was in fact a robot reacted with surprise, discomfort, fear, fascination, and repulsion:

"It's just, you see, that you don't look like a robot," said Baley, desperately.
"And that disturbs you?"
"It shouldn't, I suppose, Da – Daneel. Are they all like you in your world?"
"There are individual differences, Elijah, as with men."
"Our own robots… Well, you can tell they're robots, you understand. You look like a Spacer."[101]
"Oh, I see. You expected a rather crude model and were surprised. Yet it is only logical that our people use a robot of pronounced humanoid characteristics in this cause if we expect to avoid unpleasantness. Is that not so?" (Asimov 1997: 26f.).

Jessie lifted her eyes to R. Daneel's face, staring at it earnestly. (…) "I think you are a robot, Daneel."
And R. Daneel replied, in a voice as calm as ever, "I am." (…)

[100] In fact, Turkle found in her study of children's interactions with social robots the prominent idea among participants that "people are special because of their imperfections" (Turkle 2006a) in comparison to robots, which were regarded as superior and therefore could not become "best friends" with humans.

[101] In Asimov's robot novels, the Spacers are humans whose ancestors were the first to leave Earth in order to colonize space. They live together with a large number of robot servants on the fifty Outer Worlds out in space.

Baley's voice returned to its whisper. "How did you find out? Won't you tell me?" (...)
"Can he hear us? That thing?"
"Not if we whisper."
"How do you know? Maybe he has special ears to pick up tiny sounds. Spacer robots can do all sorts of things." (...)
"Not Daneel. They made him human-type on purpose. They wanted him to be accepted as a human being, so he must have only human senses."
"How do you know?"
"If he had extra senses, there would be too much danger of his giving himself away as non-human by accident. He would do too much, know too much." (...)
"I'm scared; I'm scared clean to death." (...)
"Why, Jessie? There's nothing to be worried about. He's harmless. I swear he is."
"Can't you get rid of him, Lije?" (Asimov 1997: 60–62).

"Dear me," said Dr. Gerrigel, with what was almost a sob in his voice, "you are a robot."
"It took you a long time to realize that," said Baley, dryly.
"I wasn't expecting it. I never saw one like this. Outer World manufacture?"
"Yes," said Baley.
"It's obvious now. The way he holds himself. The manner of his speaking. It is not a perfect imitation, Mr. Baley."
"It's pretty good though, isn't it?"
"Oh, it's marvelous. I doubt that anyone could recognize the imposture at sight. I am very grateful to you for having me brought face to face with him. May I examine him?" The roboticist was on his feet, eager. (Asimov 1997: 135).

Clousarr yelled, "Keep that thing off me."
"That's no way to speak," said Baley with equanimity. "The man's my partner."
"You mean he's a damned robot," shrieked Clousarr. (...)
Baley said, "All right, smart boy. What makes you think Daneel's a robot?"
"Anyone can tell!" (Asimov 1997: 168).

While the fictional R. Daneel could deceive most casual observers, the 'imposture' of a "realistically humanoid" robot (i.e. an *android*[102]) built with state-of-the-art technology (Fong et al. 2003; Ishiguro 2005) is much more easily recognizable but not less disturbing for people. One of several examples is 'EveR-2 Muse' (shown in Figure 37), the world's first "robot entertainer" ([INT 76]; Rötzer 2006), which was built by Baeg Moon-hong and his team at the Korea Institute of Industrial Technology (KITECH) and made its debut appearance at the Robot World Conference in Seoul in October 2006. EveR-2 Muse is a fully embodied humanoid robot that is designed to portray an attractive Korean female in her early twenties. The robot is a good five feet tall and weighs 60 kg. Its whole body is covered by an artificial skin made of synthetic pliable silicone jelly. The body has 60 degrees of freedom (DOFs), i.e. it can perform 60 independent movements with its eyes, face, neck, arms, hands, torso, and lower body. Thus equipped, EveR-2 can portray various facial expressions (including emotions like joy, anger, sorrow, boredom, etc.) and even a number of dance moves. However, it can only sit down and stand up, not walk, because the necessary mechanical components would make its legs too big for a robot that is intended to look like a human being ([INT 77]; [INT 78]). In addition, EveR-2 uses speech recognition and speech synthesis (including lip synchronization), the ability to perform hand gestures, computer vision, and other technologies in its performances.

[102] An *android* is a robot created in the human image with respect to appearance and behavior.

Figure 37. Human and machine performing together at the 2006 Robot World Conference ([INT 79]). The reader is invited to decide for himself or herself who is who.

Being a robot for entertainment, part of EveR-2's public demonstration was the singing of a song, although at its debut, it was unable to finish its song due to a technical defect ([INT 80]).

The EveR-2 robot was designed to look "like a real, flesh and blood human being" ([INT 77]), but it was widely rejected by the public. Commentators disliked its face ([INT 80]),[103] its "glazed, nightmarish glare" ([INT 81]), and it having "the bearing and presentation of an embalmed corpse" ([INT 81]). A number of (anonymized) comments on an entry about EveR-2 on the Wired Blog Network from 3 January 2007 ([INT 81]) illustrate how uncomfortable many[104] people are with things that manage to look only "almost human:"

> *I feel bad for those researchers having done all that work to have that as a result – that thing looks freaky.*
>
> *The characteristic of a 'life-like' robot that makes it creepy or freaky is its resemblance to an actual human body. Stick figures are not nightmarish because they are merely a suggestion of human form in its most basic components (legs, arms, torso), but do not closely imitate the actual appearance of a human. In other words, something only becomes scary when it is a replica which closely resembles the subject they intend to imitate. Ergo, the fact that this robot is 'freaky' and 'nightmarish' is a testament to how closely they have reproduced the features of a human body. That scariness will be removed when 'looks almost like' becomes 'looks exactly like.'*

[103] In reaction to criticism of the robot's appearance, Baeg Moon-hong said in a recent interview that his team would work to replace EveR-2's skin with a better one and maybe change its facial skeleton as well. He predicted that EveR-2 would be a 'beauty' after the successful "plastic surgery" ([INT 82]).

[104] There were also several positive (and some suggestive) comments about the robot.

(...) I think the eyes are an issue more than anything else physical. Based on the eyes, we sense the mind (or mindlessness) behind them. Eyes that don't quite look at you and don't move right, etc. will always be freaky.

(...) I'd rather have human-like responsiveness vs. semi-human looks.

(...) Talk about totally not marketable. The more human an android gets, the creepier it is. Androids should be human-shaped but look different from humans; otherwise they are scary and offensive...

In an experiment on human-robot interaction, Minato et al. (2004) investigated human reactions to the appearance and behavior of an android girl named 'Repliee R1,' by studying subjects' eye gaze behavior. The subjects had brief conversations with a human girl; the android performing eye, mouth, and neck movements; and the motionless android. Minato et al. found that subjects looked at the eyes of the android more frequently than at those of the human girl, despite the general tendency in Japanese culture to avoid eye contact. Many subjects commented on the artificiality of the android's appearance and behavior (especially its eye movements), as well as on an imbalance (cf. Chapter 7.3) between the two aspects. Hence, this experiment provides some evidence that people tend to show negative reactions to robots which try to, but fall short of, appearing and behaving completely human. "Illusions [of human-likeness], to be believed, must not only look, but act 'real'" (Tognazzini 1993).

All these examples from movies, the science-fiction literature, and robotics indicate that the relationship between the increasing realism of animated characters or robots and the emotional response of humans toward them may not be linear (Dautenhahn 2004: 51), in the sense that more realistic is automatically better. In 1970, the Japanese robot scientist Masahiro Mori (1970; 1982) proposed the principle of the *uncanny valley* for life-like robots (cf. Figure 38), which more recently has also been applied to computer-animated characters in movies (Ebert 2004) and computers (e.g. Dautenhahn 2004; Romano 2007; Valvoda 2007; Aylett et al. 2008). Mori's principle predicts that as a robot is made more human-like in terms of appearance and movement, it becomes more and more familiar (or believable) to humans, resulting in an increasingly positive human emotional response, until a local minimum is reached at which the response suddenly becomes strongly repulsive. At this minimum (the uncanny valley), the robot's design blurs the line between human and non-human and the robot might be mistaken for real (as R. Daneel in Asimov's novels); however, slight but noticeable differences can quickly destroy this impression and create a disturbing feeling in humans that results in rapidly declining approval rates. Dautenhahn calls this experience the *Zombie effect* (Dautenhahn 2004: 51). An uncanny robot or agent is like one of the living dead: it looks human-like to observers until they examine it more closely and are horrified by the discovery that it is not human. As Tognazzini put it, "[t]he better the illusion, the more shocking is any flaw" (Tognazzini 1993). In the context of facial animation, Parke wrote earlier:

The quest for realism runs headlong into the fact that a 'realistic' face which is just a little wrong, receives strong critical reaction from most viewers. (Parke 1991: 229).

Sundar (1994) reported the same effect for computers that are made more and more human-like to facilitate the interaction with the user: at a certain point, users reject these "almost human" computers because they fail to live up to the heightened expectations (of human-like interaction capabilities) which they raise. The problem is that "[a] human being will always suppose that, the more human a robot [or computer] is, the more advanced, complicated, and intelligent he [sic!] will be" (Asimov 1994: 428). In sum, it appears that the closer designers

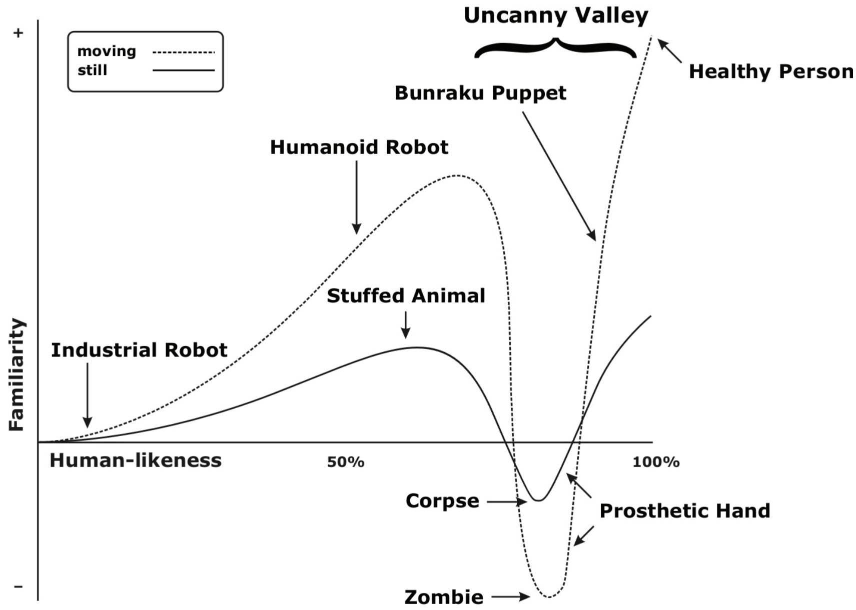

Figure 38. Masahiro Mori's uncanny valley (Mori 1970). Adapted from MacDorman (2005, Figure 1). The diagram plots the familiarity (believability) of a robot against its human-likeness (realism). The uncanny valley is the sudden dip in emotional response of humans to near-human robots. As the dotted curve shows, the emotional response is amplified by movement.

come to achieving human-likeness, the closer they move to the abyss and the more likely their design is to fail.

However, Mori's model predicts that it is possible to escape from the uncanny valley by increasing the anthropomorphism (cf. Chapter 9.1) of the robot's appearance and movement until the robot is truly indistinguishable from a healthy human being (100% similarity) and the human emotional response has reached its maximum after another steep rise.[105] It should be noted, though, that the rightmost part of the chart (i.e. the rise out of the uncanny valley by making robots completely human-like) is problematic because Mori had no basis for that prediction when he developed the chart more than thirty years ago, as human-like robots are technologically possible only today (and still only partially).

The general lesson of the uncanny-valley model for designers of embodied interactive agents is one of self-limitation, i.e. to aim for less than they might be able to accomplish in terms of realism with the current state of technology, which is not sophisticated enough to create 100%

[105] At this point, one can only speculate about what might happen after this second maximum, as embodied agents are improved even further and eventually become 'super-human.' Humans are not perfect, neither in appearance nor in behavior. If embodied agents become superior to humans in one or both of these aspects some day, how will people react? In the case of R. Daneel, the very perfection of its physical appearance betrayed its non-humanness, amplified by the contrasting imperfection of its behavior. Overall, people's reactions to R. Daneel were not favorable. In Turkle's study (2006a), children regarded social robots as superior and therefore thought that they could not become "best friends" to humans, who they regarded as special because of their imperfections. This might indicate that humans possibly will not appreciate such 'super-human' agents.

human-likeness of artificial agents.[106] Stepping down from the goal of highly realistic agent representations is necessary because

> *both research and experience suggest that a mismatch between realism in appearance and the apparent knowledge level of the agent can have a deleterious effect on credibility [(cf. Chapter 7.3)]. In other words, the more visually realistic the representation, the higher the expectations of the user in relation to the appropriateness and 'intelligence' of utterances and actions. Agents that 'look' smart and 'act' or 'talk' dumb are poorly perceived by many users, who express* ***a higher tolerance for the limitations of a 'character' more sketchily represented****, for instance through cartoon-like graphics. (Dowling 2002, emphasis added).*

To prevent their agents from falling into the abyss of the uncanny valley (very human-like but still falling short), designers should content themselves with reaching the first peak in Figure 38 (human-like but less ambitious) rather than the second (completely human-like but unattainable) (MacDorman 2005). Human emotional responses reach their first local maximum at the first peak, which may be a sufficient criterion of success for many applications involving embodied interactive agents.

Arguments for aiming at less than photorealism in agent design also come from McCloud's work on how people compose, read, and understand comics (McCloud 1993). According to McCloud, simplifying and reducing representations of characters, making them iconic rather than realistic, can increase audience involvement because the mental concept and image which the reader has of *himself or herself* is highly iconic (in contrast to his or her mental image of other people). Therefore, users may find it easier to relate to embodied agents with an iconic (simplified and reduced) representation. Thus, these agents can make a stronger impact on users (Gulz & Haake 2006: 330). However, for any given agent, the appropriate design with respect to the iconic-realistic dimension depends on both the target group of users (cf. Chapter 6.1) and the agent's role (cf. Chapter 7.14). Different cultures and subcultures may respond in different ways to agents that are more realistic vs. more iconic. Furthermore, for a pedagogical agent in the role of tutor, a higher level of realism may be appropriate because, playing a different role than the learner, it portrays "the other" more than an agent acting as a learning companion (a role closer to the learner) (p. 331).

While the first peak in Figure 38 might represent an appropriate goal level for agent realism, the question arises whether there is also a low-end level of human-likeness that is just enough to make people believe that the illusion portrayed by an agent is real. Tognazzini (1993) called this point the *threshold of believability*. He suggested that with careful design and attention to detail, a point or threshold can be reached beyond which disbelief turns into belief in the reality of the illusion that has been created. Since the exact threshold is subjective, depending on the person and his or her mood, designers "must exceed it sufficiently to ensure believability" (Tognazzini 1993). A similar idea was mentioned by Nowak (2004). She speculated that "there may be a *sufficiency threshold* where an intelligent other begins to be perceived as intelligent, or human enough, for the purposes of interaction. (…) [I]f there is a sufficiency threshold, it is

[106] However, there is optimism among developers that they will be able to create artifacts which are indistinguishable from humans soon. For example, in an interview with the BBC published online on 20 February 2007 ([INT 83]), a producer of computer games admitted that "games are currently in an 'uncanny valley.' They look strange – they're too close to real, but not quite real." However, he added: "Give us another year or two, and we'll be able to completely get across that uncanny valley." On the other side of the valley, developers hope to achieve 'super-realism,' i.e. (game) characters that demonstrate all the naturalness and subtleties of human appearance and body language which they currently still lack ([INT 83]).

likely that the standard for social interactions would be higher than the standard for task-based interactions." (Nowak 2004, emphasis added).

The challenge for agent design is to create the right level of realism. Both Mori's model, which suggests that movement amplifies the impact on the user (cf. Figure 38), and empirical studies indicate that behavior in context seems to be more important than appearance (Dautenhahn 2004: 52f.).[107] As a result, not only the agent's embodiment but importantly also its behaviors (speech, locomotion, gestures, and facial expressions) require an appropriate degree of realism in order to enhance believability and support the agent's tasks.[108] In general, a body should exhibit the behaviors which users expect from this kind of embodiment, observing the relevant restrictions. For example, exaggerated facial expressions or gesturing may be acceptable for a cartoon figure but would undermine users' perception of an agent with a photorealistic human-like appearance (McBreen & Jack 2001: 404). Other research found that the more anthropomorphically-realistic a character becomes, the more human-like it is expected to act (Vinayagamoorthy et al. 2005).

Agent embodiments can be rendered as two-dimensional (2D) or three-dimensional (3D) representations (Churchill et al. 2000: 77). The 2D graphics are easier to draw and process and allow greater levels of detail. The 3D models can be viewed from any angle and can be created and animated on the fly using well-defined rules. McBreen and Jack (2000; 2001) found that the participants in their study preferred the 3D agents to 2D agents and the 2D and 3D fully embodied agents to 2D and 3D talking heads.

Whether an embodied agent is designed as a stick figure or a more realistic image, in 2D or in 3D, is a trade-off between task and context parameters, the constraints of rendering (Churchill et al. 2000: 77), and the need to avoid the pitfall of the uncanny valley. Portraying an agent with rough strokes might be sufficient to make a first impression, but the finer details are necessary to complete the picture, and their presence (or absence) becomes more and more apparent with increasing exposure of users to the agent (Loyall & Bates 1997: 106; Shaw et al. 1999: 288).

---

[107] Cassell and Tartaro (2007) made essentially the same point by arguing that rather than attempting to impress the user through the agent's realistic or life-like appearance, agent designers should seek to achieve *intersubjectivity* in human-agent interaction by making their agent *behaviorally* realistic, in particular with respect to its verbal and non-verbal communicative behaviors (cf. Chapter 10.1 and Chapter 11), thus inducing the user to unconsciously respond to the agent as if it was like a human partner in the conversation. Along the same line, experimental results of Iacobelli and Cassell (2007) suggest that perceptions of agent ethnicity can be influenced by changing verbal and non-verbal behaviors of an embodied conversational agent (cf. Chapter 11.3.4) rather than its physical traits.

[108] Bailenson et al. (2005) examined how assessments of copresence (or social presence) with embodied agents in immersive virtual reality (measured in terms of both participants' perceptions of and responses to the agents) were influenced by the degree to which each agent resembled a human being in appearance and behavior. Copresence was found to be lowest when the appearance and behavioral realism of an embodied agent were highly inconsistent (cf. Chapter 7.5) with each other. Similarly, Garau et al. (2003) demonstrated that increasing the photorealism (i.e. the appearance) of an avatar (the virtual representation of a human being, cf. Chapter 3.1.4) without also improving its behavioral realism (they investigated eye gaze behavior) can reduce the perceived quality of communication (operationalized in terms of copresence, among other indicators).

9.6 Complete vs. Partial Bodies

Agent embodiments do not always have to be complete (humanoid) bodies but may range on the completeness scale from just a head to a head and neck to a head, neck, arms, and torso to a fully articulate body. In general, partial embodiments are easier to create and animate but have a limited potential for action and interaction.

The extent of completeness required for an agent embodiment depends on the expressive, gestural, and locomotive capabilities necessary to support the agent's role, persona, and relationship-building efforts (cf. Chapter 8.4). *Full bodies* are suitable for agents whose role involves moving around in the environment and gesturing at or manipulating things. If the agent is meant to remain stationary (e.g. appear in a window), it can be enough to model the *upper half* of the body. However, designers should make sure that both the face and the hands are visible since they are important communicative devices. To focus the user's attention on the head region, the agent's embodiment can be restricted to a *talking head*. This focus can help users to better understand the agent's verbal messages because auditory and visual speech contain complementary information (Massaro 1998, cf. Chapter 10.1.3). In language learning and speech training, talking heads can become transparent for a physiologically accurate and pedagogically useful display of speech production (Massaro et al. 2000; Massaro 2005).

9.7 Stability

Agents may not remain in the same body all the time but change their representation during the course of an interaction or across a number of interactions (Churchill et al. 2000: 75). Reasons for an agent to choose a different embodiment include:

- *Changes in the environment*. For example, if the user switches from a regular computer to a mobile device, the agent may have to change to a simpler embodiment (cf. Chapter 14.3).
- *Changes in task requirements*. For example, an agent for language training can switch from a full body to a talking head in order to illustrate articulatory patterns (cf. Section 9.6).
- *Changes in user preferences*. The user may like a different embodiment better than the current one.

The possibility of an agent to change its embodiment over time is at odds with the design principle of consistency (cf. Chapter 7.5) because variable embodiments contribute to different overall impressions, making it difficult for the agent to maintain its identity. It may also not be what users prefer because people in the real world generally do not change from a human being to a bouncing ball to a paperclip to whatever during or across interactions.

No matter what embodiment an agent may currently assume, designers must ensure that the chosen body does not block the user's view or access to applications and information. There are even contexts where a voice without a body may be preferable (Churchill et al. 2000: 75).

9.8 Summary

This chapter was concerned with the design of the digital representation through which an embodied interactive agent appears in and acts on its environment. The embodiment has to be designed with care because it is the first characteristic of the agent that is perceived by users, indicating the agent's availability for interaction, its traits, capabilities, and limitations. Crafting an appropriate embodiment for an agent requires designers to make choices from a multi-dimensional space of design options, which were discussed in separate sections of the chapter.

The first dimension (cf. Section 9.1) concerns the degree to which an embodied agent should resemble a human being in appearance. Arguments for and against using anthropomorphic forms in user interfaces were presented, followed by an elaboration of the author's own view in favor of a carefully considered use of human-like features in ways that serve a function beyond entertainment and render the interaction multimodal.

The design of the appearance of embodied interactive agents was the topic of Section 9.2. The discussion stressed the importance of appearance for agents to make a (preferably positive) first impression on users, as well as its impact on how effective and credible the agent appears in its role and on the ways users interact with it. A list of guidelines for designing agents that appeal to users from different cultures was presented.

Embodied agents should not only appear attractive but also individual, i.e. they should reveal and express a unique and distinctive persona in every possible way, which includes their embodiment. In Section 9.3, it was argued that the design of an individual embodiment should be based on an intimate knowledge of the character, which is captured in a character profile.

Agent embodiments differ with respect to their capability of movement (cf. Section 9.4), with design options ranging from static images to fully animated, expressive figures. Movement is a prerequisite for embodied communication and can support the agent's role or enhance its persona.

Embodied agents can be portrayed with different degrees of realism and with different depths of the images portraying the agent. The realism dimension was discussed in detail in Section 9.5. Using Mori's principle of the uncanny valley, the case was made against creating agent images that portray virtual humans as (photo-) realistically as possible because such images tend to provoke strong negative reactions from observers if they are not completely indistinguishable from images of real people.

The embodiments of agents are more or less complete (cf. Section 9.6). The decision to use a representation consisting of the head, the head and neck, the upper half of the body, or a fully articulate body involves a trade-off between the effort required to create and animate the embodiment and the range of actions and interactions that have to be supported.

The final aspect of agent embodiments that was covered in the chapter is the stability of the agent's representation (cf. Section 9.7). Agents may switch to a different body to accommodate changes in the environment, task requirements, or user preferences, but it was argued that doing so is at odds with the design principle of consistency (cf. Chapter 7.5) and with the preferences of users.

10 Behavior

Behavior is a collective term for the actions and reactions of humans, animals, and inanimate entities in a given situation, usually in relation to their environment which includes other animate and inanimate entities of the same or a different kind. All behaviors involve to some extent the use of a natural or artificial body with certain capabilities and limitations (cf. Chapter 9), especially those behaviors performed to produce an observable and interpretable effect on the environment through physical movement. Particular behaviors can be described in terms of multiple dimensions (cf. Figure 39):

- *Unconscious .. conscious*. Unaware vs. deliberate and intentional behavior;
- *Covert .. overt*. Hidden vs. observable behavior;
- *Involuntary .. voluntary*. Uncontrollable vs. controlled behavior;
- *Simple .. complex*. Elementary (or atomic) vs. composite behavior;
- *Unacceptable .. acceptable*. (Not) satisfactory or allowable behavior;
- *Innate .. learned*. 'Built-in' vs. acquired behavior;
- *Stereotypic .. individual*. Conforming vs. idiosyncratic behavior;
- *Non-human .. humanoid*. Behavior that is (un-) characteristic of the human species;
- *Unrealistic .. realistic*. Behavior that does (not) observe physical restrictions.

Behaviors are influenced by many factors. In the case of human behavior, these factors include:

- *Culture*. The set of distinctive spiritual, material, intellectual, and emotional features of society or a social group ([INT 52]), "a theory for interpreting the world and knowing how to behave" (Gudykunst & Kim 1992);
- *Attitude*. A complex long-term mental state that has semantic content and is directed at particular instances or whole categories of events, individuals, objects, or situations and is reflected in a person's behavior;
- *Emotional state*. The current internal dynamics of an individual (including aspects of both mental and physical state) in response to internal and/or external stimuli deemed relevant to his or her needs, goals, or concerns (cf. Chapter 13.1);
- *Personality*. The complex of psychological traits that uniquely characterize an individual (Hayes-Roth et al. 1997, cf. Chapter 13.5);
- *Values*. Individual principles and standards of behavior and one's priorities in life;
- *Ethics*. Moral principles that govern a person's behavior or the conducting of an activity, helping to distinguish right from wrong;
- *Authority*. The power or right (of others or oneself) to give orders, make decisions, and enforce obedience;
- *Rapport*. A relation of mutual understanding or trust and agreement with others;
- *Hypnosis*. An induced state of consciousness in which an individual is highly susceptible to suggestions or directions from another person that alter his or her perception, thinking, and behavior;

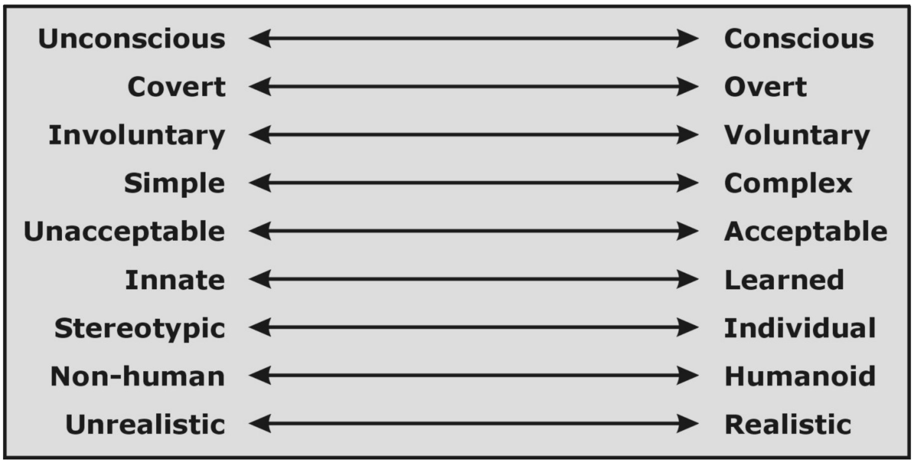

Figure 39. Dimensions of behavior.

- *Persuasion*. (1) The intentional effort to change someone's attitudes or behaviors or both, without using coercion or deception (Fogg 2003: 15);[109] (2) a personal belief or judgment of an individual that is not founded on proof or certainty;
- *Coercion*. Exertion of force or threats to make an unwilling person do something.

Social behavior is meaningful behavior that takes place in a social context and results from the interaction between and among individuals. The purpose of many social behaviors is communication. Humans can use language and the special characteristics of their bodies to engage in conversations with one another (Cassell 2000b). Furthermore, social behaviors serve to sustain the image of a person as a unique individual, and they are instrumental in building social relationships with other people (cf. Chapter 7.16). Likewise, embodied agents that exhibit affordances (cf. Chapter 6.1) similar to those of the human body can emulate types of human social behavior that facilitate communication with the user, enhance the agent's believability, and help to establish human-agent relationships. These three aspects are discussed in Section 10.1, Section 10.2, and Section 10.3, respectively. Section 10.4 is concerned with the level of presence (high or low profile) that should be maintained by embodied interactive agents in their environment. The chapter concludes with a summary in Section 10.5.

10.1 Communication

Agents with the ability to speak and virtual bodies that integrate some or all of the central elements of a humanoid embodiment (including a face, hands, feet, and a torso, cf. Chapter 9.1) can be designed to keep up their side of a multimodal dialogue with the user, which involves displaying interdependent verbal and non-verbal behaviors. An embodied conversational agent

[109] Not only people have the power to induce changes of attitudes and behaviors in other people but also technology. Computers as persuasive technologies are the focus of *captology*, an emerging field of research concerned with the "design, research, and analysis of interactive computing products created for the purpose of changing people's attitudes or behaviors" (Fogg 2003: 5). Given that education at its heart is about changing the attitudes and/or behaviors of those being educated, this research becomes immediately relevant for the design of e-learning technologies in general and pedagogical software agents in particular.

(cf. Chapter 11.3.4) has the ability to coordinate verbal and non-verbal means (eyes, face, hands, voice, posture, and movement) during its interactions with the user in order to (Cassell 2000b; Cassell 2001):

- *Regulate* the conversation process (e.g. contrast, emphasis, feedback, and turn-taking);
- *Contribute content* to the ongoing conversation (e.g. use of the hands to indicate the size, distance, position, etc. of something referred to verbally);
- Convey *emotional and cognitive states* (e.g. confusion, sadness, or thinking).

Although face-to-face human-human interaction is rich in non-verbal communicative behaviors, until recently, only emblematic gestures (cf. Section 10.1.2) and emotional facial displays (cf. Section 10.1.3) have been integrated to computer interfaces (Cassell 2000a: 5). However, Cassell and Thórisson argued that:

> *The importance of embodiment in computer interfaces lies first and foremost in its power as a unifying concept for representing the processes and behaviors surrounding conversation. (Cassell & Thórisson 1999: 521).*

In other words, emulating human non-verbal communicative behavior may be more important for an embodied agent than providing emotional feedback because it allows users to apply their knowledge about how to participate in human-human conversation and gives them confidence that the interaction is proceeding smoothly since the agent's non-verbal behaviors keep them updated about the process of conversation throughout the interaction. Cassell and Thórisson found in their study that agents showing non-verbal behavior related to the process of conversation were rated as more helpful, more life-like, and smoother in their interaction style than agents lacking non-verbal behaviors or agents giving emotional feedback and that users interacting with agents exhibiting non-verbal cues required fewer utterances to accomplish the task.

The communicative content portrayed by an embodied agent is a function of all the behavior channels which its embodiment provides and which are used in the interaction, including both verbal and non-verbal channels. Users' perception of the agent's internal state depends on the content of each of these channels and their relationships. For the design of embodied interactive agents, non-verbal communicative behaviors can be conveniently grouped into the categories of gesture, face, gaze, posture, and locomotion, to be discussed below. The design of a voice for an embodied interactive software agent is the topic of the next section.

10.1.1 Voice

Speech is behavior that creates linguistic acts by means of voice to communicate messages from a sender to a recipient. *Voice* is sound that is characterized by features including pitch, intensity, quality, and condition. Humans produce voice by modifying and modulating an incoming or outgoing air stream with the vocal cords (vibration) and the articulators of the vocal tract. People convey rich messages not only through the content of their speech (cf. Chapter 7.15.4) but also through their manner of speaking (cf. Chapter 7.15.5), i.e. through how they say things (Isbister 2006: 183). The voice conveys *paralinguistic cues* which provide information over and above the words of the spoken message. Some of these cues arise from the physical qualities of a speaker or their body's involuntary reactions to circumstances (p. 184). For one, voice conveys cues that automatically and efficiently indicate the identity of an individual. Listeners use the pitch of a voice to make assumptions about the age, biological sex, and other aspects of the speaker. Pitch range, loudness, and range of loudness can also help listeners to identify people, but like pitch, these aspects may change depending on the situation. In addition, listeners analyze the timbre of voices and use this information to describe a voice as smoky, nasal, breathy, raspy, shrill, etc. (Nass & Brave 2005: 98). *Accent* (i.e. the characteristics of pronunciation inherent in the speech of a particular group of people) is present in the voice of any speaker and, along with word choice, serves as an important indicator of place of origin (Nass et al. 2006: 380): "everyone sounds like they were brought up somewhere" (Nass & Brave 2005: 62). A person's voice can also betray mood and emotion since vocal production is affected by the physiological responses associated with affective states (cf. Chapter 13.1.1). Furthermore, a person can alter the prosodic features (i.e. pitch, intensity, and timing) of his or her speech (cf. Chapter 11.1.2.1) to adopt a pacifying, pleading, arrogant, or neutral tone of voice. Different emotions have their own signatures in terms of vocal cues (cf. Chapter 13.4.2). The vocal style is also influenced by the speaker's personality (cf. Chapter 13.5). For example, the voice of a dominant person exhibits low pitch, limited pitch range, and loudness (Nass & Brave 2005: 43). Personality is also a predictor of language style. Dominant speakers construct shorter utterances but still take more turns and set both the tone and the topics of a conversation. In contrast, submissiveness is signaled by a softer voice at a higher pitch, longer utterances, and following the other's lead in conversation (Isbister 2006: 31f. + 188). The same speaker can adopt different voice (and language) patterns depending on his or her addressees (friends, children, teachers, superiors, subordinates, etc.). In other words, voice also reflects social roles and relationships (p. 188).

Voice has been consistently identified as a central feature for making an agent seem person-like (Reeves & Nass 1998; Moreno et al. 2001; Atkinson 2002). However, the positive effects of voice extend beyond believability to include user motivation and performance. For pedagogical software agents, voice was found to strongly indicate a social presence, resulting in increased learner interest and interaction with the agent (Baylor & Ryu 2003). Nass and Lee (2001) presented both extrovert and introvert people with unambiguously computer-generated voices (extrovert or introvert) reading book reviews and found that their subjects not only accurately recognized personality cues in the computer-synthesized speech but also showed similarity attraction in their evaluation of the computer voice, the book reviews, and the reviewers, preferring the voice, reviews, and reviewers they perceived as matching their own personality. In a later study, Lee and Nass (2005) demonstrated that users feel stronger social presence (cf. Chapter 5.4.5) in e-commerce contexts (cf. Chapter 3.2.2) when they hear a computer-generated voice that manifests a personality which is similar to the user (especially to the extrovert user), is consistent with the personality conveyed by the phrasing of descriptions, and is extroverted rather than introverted.

With respect to the impact of accent, Nass and Brave (2005: 65ff.) described an experiment in which participants socially identified with agents based on the cultural background that was evident in the paralinguistic cues of accent manifest in the voice of an agent but not based on the racial background evident in the photograph of an agent (i.e. the agent's appearance). They concluded that "accent, which reflects the culture in which the person grew up, is critical for responses to interfaces, even when knowledge of an agent's (or person's) ancestry is readily apparent" (Nass & Brave 2005: 67). In addition to social identification, the accents in agents' voices may also elicit stereotypes (about degrees of intelligence, elitism vs. egalitarianism, etc.), much in the same way as gender of voice (p. 68). Furthermore, the participants in the study preferred consistency between the race of an agent portrayed in its photograph and the accent of the agent's voice. They felt disturbed when there was a discrepancy between the way the agent looked and the way it sounded. Therefore, "[t]he race of the face should 'match' the accent of the voice" (Nass & Brave 2005: 139).

Regarding performance of learners, Moreno and Mayer (2000) found that students who communicated with a pedagogical agent via speech rated a multimedia science lesson more favorably, recalled more, and showed better ability to apply what they had learned in problem solving than students who learned the same verbal materials as on-screen text. Along the same line, experimental results obtained by Moreno et al. (2001) indicate that a pedagogical agent coaching via spoken narration rather than on-screen text when students are learning from visual presentations can promote meaningful learning in multimedia lessons because the dual mode of presenting information aurally and visually makes more effective use of working memory and creates deeper understanding than the combination of verbal and non-verbal visual materials (Moreno & Mayer 1999: 366). Hence, the ability to communicate via speech is essential for an effective pedagogical agent, both to convey its persona and to support its instructional interactions (cf. Chapter 12.3).

For any speech-enabled application, a voice has to be chosen that meets at least two requirements: intelligibility and naturalness. The voice is *intelligible* if human listeners can understand the sounds and words produced, even under bad conditions (e.g. noise). It sounds *natural* if it as easy to listen to as a human voice, in particular when it is used to deliver longer messages. This does not imply that the voice has to sound like a real human voice.

In addition to clarity and tonal quality, *expressiveness* is another criterion, i.e. the extent to which a voice exhibits cues (intensity, pitch, pitch range, speech rate, etc.) that indicate the speaker's emotional, mental, and physical state; personality; attitudes; and relationship with the listeners. The voice selected for an embodied interactive agent should consistently convey the agent's personality and its current emotional state. People have a preference for consistency between the emotional tone carried by the voice and the emotional information implicitly conveyed by the words of a message, the only exception being clear cases of sarcasm. Inconsistency between verbal and vocal emotions is interpreted as a sign of insincerity, instability, or deception (Nass & Brave 2005: 86). In general, designers should take care to ensure consistency of the voice with the behaviors, attitude, and language of the agent to which it will be attached (Reeves & Nass 1998: 177). Since voices also activate stereotypes associated with age, ethnicity, gender, personality, etc. (Nass et al. 2000a), selecting a voice that is acceptable for the target audience becomes even more important.

Designers can realize the voice of an agent using a human voice talent, a voice produced by a speech synthesizer (cf. Chapter 11.1.2), or a combination of both. Using a human voice in an agent is the most expressive and natural sounding but also the least flexible solution because all utterances have to be recorded beforehand. Additions or changes to the agent's messages make further recordings necessary. In addition, voice design becomes dependent on the availability of a particular speaker. As a result, the ability of agents with recorded human voices to participate flexibly in (open-ended) instructional conversations with the learner is obviously limited. Recorded human speech is thus an option for embodied interactive software agents which:

- Require high-quality speech output (e.g. in foreign language learning);
- Deliver only a fixed number of messages;
- Do not have to participate in open-ended conversations with the learner.

Synthetic speech (Dutoit 1997; Holmes & Holmes 2001; Dutoit & Stylianou 2003; Tatham & Morton 2005) is generated by computers either from text (text-to-speech) or from an internal representation of information (concept-to-speech). These two approaches will be reviewed in

Chapter 11.1.2. Using speech synthesis in embodied interactive agents has the following advantages:

- The agent's messages can be generated dynamically.
- The output of a speech synthesizer can be controlled to realize different voices or speaking styles.
- State-of-the-art speech synthesis systems can produce highly intelligible output, which sounds quite natural for short, simple utterances.

However, the ability of current speech synthesis technology to produce natural-sounding and expressive speech is limited (see Chapter 11.1.2.4 for a discussion of the state of the art in speech synthesis). Currently, synthetic speech is an option for agents which:

- Must deliver messages of unpredictable number and content;
- Do not require speech output of the highest quality;
- Can get by with default (neutral) prosody;
- Can keep their contributions short.

While from a developer's perspective, the advantages and disadvantages of recorded vs. synthetic speech and the resulting guidelines for their application are clear, it is also necessary to consider the preferences of users. Here, the experimental evidence suggests that users prefer natural-sounding human voices. For example, based on the results of two experiments involving learners listening to the same narrated animation delivered by a human voice with a standard accent, a human voice with a foreign accent, and a machine-generated voice, Mayer et al. formulated a *voice principle* for the design of multimedia instructional messages, which states that "students learn more deeply when the narration in a multimedia lesson is spoken by a standard-accented human voice rather than a foreign-accented human voice or a machine voice" (Mayer et al. 2003: 424, cf. Table 1). The findings of Mayer et al. were later replicated and extended by Atkinson et al. (2005). Along the same line, Huang et al. (2000) found that an interface using recorded human speech was perceived as more useful and enjoyable than the same interface using a synthetic voice (although the performance of the interface was identical in both conditions) because the synthetic speech, while fully intelligible, was less fluent and more difficult to listen to than natural speech, which made the interface seem less competent and enjoyable than the recorded-speech version (Nass & Brave 2005: 120). However, the participants in Huang et al.'s study did not have an absolute preference for 'human-like' or 'machine-like' interfaces; rather, consistency between voice and use of first-person language was important. The study investigated the question when voice user interfaces should use the first person singular and when they should avoid the use of 'I,' in other words, under what conditions they should position themselves as human or non-human. The experiment used recorded-speech and synthetic-speech voices that did or did not use the pronoun 'I.' The results suggested that recorded-speech voice user interfaces should say 'I,' whereas synthetic-speech voice user interfaces should not say 'I' (Nass & Brave 2005: 119), although both can be regarded as social actors because they receive some human attributions (p. 123). Interfaces with a synthetic voice were felt to be not human enough to use the language of personhood, due to the perceived mismatch between the non-human voice of the machine and the use of 'I.' In the words of Nass and Brave, "it is not acceptable for non-humans to say 'I'" (Nass & Brave 2005: 120). Huang et al. (2000) recommended that *designers should pair human-like voices with human-like, personal scripts (after all, most humans use "I" and "me" quite often), while they should combine machine-like voices (i.e., synthesized speech) with less human-like, passive scripts (Huang et al. 2000: 4).*

Two things are worth noting about this experiment. First, the interface was voice-only and did not involve an on-screen character with a face and a body. Other experimental studies used talking agents with faces and voices. These studies showed that participants preferred consistent face-voice combinations (human face/human voice or synthetic face/synthetic voice) over interfaces mixing synthetic and human faces and voices (see Section 10.1.3 for a detailed discussion), thus lending further empirical support to the design principle of consistency (cf. Chapter 7.5). Second, due to continuing efforts to improve speech synthesis technology (cf. Chapter 11.1.2), the gap between natural human voices and synthetic computer-generated voices is closing. Therefore, it is possible that synthetic voices may sound human-like enough to deserve the 'right' to deliver first-person language in the near future.

Mullennix et al. (2003) conducted an experiment which addressed the question whether gender of voice and gender of listener affect the social perception of human speech and synthetic speech in a task which presents listeners with a persuasive argument delivered in either male or female human or synthetic speech. Participants preferred female human speech to female synthetic speech and male synthetic speech to female synthetic speech. (Two previous studies (Stern et al. 1999; Stern et al. 2002) found that a male human voice presenting a persuasive argument was rated more favorably than a male synthetic voice delivering the same argument.) Similar patterns of ratings by female and male listeners across human and synthetic speech suggested that people stereotype both human and computer-generated voices for gender in a similar way. Hence, for computer-based applications with the intent to persuade, i.e. to change people's attitudes or behaviors (Fogg 2003: 15), a human voice is the best choice, at least until synthetic speech has caught up in quality. If synthetic speech is used, a male rather than a female voice should be chosen.

A combination of human and synthetic speech is currently used in voice-based systems that deliver the fixed part of a message in recorded speech and employ speech synthesis to fill slots with variable information. However, several studies (Nass et al. 2000b; Gong & Lai 2001; Gong et al. 2001) found that homogeneity of voice (i.e. use of synthetic speech for all messages) in an interface inspired more trust, liking, and perceptions of competence than a mix of a recorded voice and a synthetic voice. This suggests that mixing a human voice and synthetic speech in the same application may be ineffective because users expect consistency in the voices they hear. Users may find it easier to concentrate on their task if they do not constantly have to adapt to a different voice. Therefore, an embodied interactive agent should use either recorded human speech or synthetic speech but not mix both:

[I]nterfaces do better when they speak uniformly in synthetic speech, even if the speech is of poorer quality, rather than shifting back and forth between synthetic and higher-quality recorded speech. (Nass & Brave 2005: 184).

10.1.2 Gesture

The term 'gesture' includes both volitional and non-volitional movements of different parts of the body, although gestures are often associated with hand movements. In communication, gestures serve to emphasize, clarify, or amplify what a participant is saying. In addition, gestures help people to regulate conversational exchanges and to express emotions (Allbeck &

Badler 2004: 111). Gestures used in face-to-face interactions include emblematic, propositional, and spontaneous gestures (Cassell 2000a: 6–11).

Emblematic gestures are stereotypical patterns with understood semantics, which are performed consciously by speakers to symbolize something abstract. For example, the good-bye wave signals 'farewell,' the fingers forming a 'V' stand for 'victory,' the stretched-out palm of the hand means 'stop,' the "thumbs up" gesture shows that things are 'good' or 'Ok,' and a nod or shake of the head indicates assent or dissent. These and other emblematic gestures are 'coded,' i.e. represented in memory with a direct link between a gesture and its meaning.

However, the interpretation of emblematic gestures is subject to cross-cultural variation. The same meaning may be expressed by different emblematic gestures in different cultures, and a given emblem may have different interpretations across cultures. Gestures indicating assent and dissent are particularly dangerous in this way. For example, the ring gesture involving the thumb and forefinger forming a circle and the remaining fingers extended will be interpreted by Americans as 'OK' but will be insulting for Germans, Russians, and a wide range of other cultures (Allbeck & Badler 2004: 112). A simple head nod can indicate agreement or attentive listening, confirm turn-taking, or express denial (Maldonado & Hayes-Roth 2004: 163). Embodied interactive agents thus need to carefully adapt their culturally coded emblematic gestures to the culture of their target audience. Other emblematic gestures are biologically coded (e.g. raised fists expressing elation, and other ritualized body movements) and might be the same across cultures (de Rosis et al. 2004: 84).

While cultures differ in the number of emblematic gestures in their communicative repertoire,[110] these gestures form only a minor portion of the gestures produced in everyday communication (Cassell 2000a: 7f.). Despite their minor role, interface designers tended to focus only on emblematic gestures at one point. For embodied agents designed to emulate face-to-face conversation, the priority should rather be to integrate those gestures that co-occur with speech in face-to-face dialogue (p. 8).

Propositional gestures are conscious gestures that complement or elaborate on speech content in a conversation, indicating direction, simultaneity, size, spatial relationships, and so on. Examples of propositional gestures include:

- Indicating the size of an object being described while saying "It was this big;"
- Pointing at an object and then pointing at a destination and saying "Put that there."

Propositional gestures are typically found in situations where the conversation is about the physical location of the exchange. However, like emblematic gestures, they are not the most frequent gestures in spontaneous conversation (Cassell 2000a: 8). Embodied agents require propositional gestures to engage in conversations with the user that deal with a shared task. Such task-oriented dialogues (cf. Chapter 3.2.1) require both parties to make reference to objects in the environment, for example when giving instructions.

The vast majority of gestures occurring in face-to-face dialogue are *spontaneous gestures*. While unconscious and unintentional (in contrast to emblematic and propositional gestures), they are important devices for conveying communicative intent when co-occurring with speech. The different types of spontaneous gestures include the following (Cassell 2000a: 8–11):

- *Iconic gestures* have a form that depicts some physical aspect of an action or event, such as the shape and size of an object or the way a described action is performed.

[110] For example, French and Italian cultures have more emblematic gestures than American culture (Kendon 1995).

- *Metaphoric gestures* represent by their form an abstract feature of the subject matter (e.g. exchange, emergence, or use). *Process metaphors* (e.g. a rolling gesture) indicate an ongoing process. The *conduit metaphor* (e.g. a box gesture accompanying the introduction of a new topic) represents an abstract idea as a bounded container that can be held and passed between interlocutors.
- *Deictic gestures* represent or locate both abstract and physical entities referred to in the current conversation in physical space (e.g. pointing to an object while saying “this one wasn’t here last time”).
- *Beat gestures* are small hand movements that occur with accented spoken words and speaker turn-taking. For example, a beat accompanying the word ‘blue’ in “I would take the [blue] pill” stresses which pill should be taken.

The manner (or *style*) in which an embodied interactive agent performs emblematic, propositional, and spontaneous gestures provides important cues as to the agent’s personality, cultural background, cognitive and emotional state, relationship to the interlocutor (e.g. boss/assistant), and the conversational setting (e.g. formal vs. informal) (Noot & Ruttkay 2005: 213).[111] The observation that an agent performs such non-verbal communicative behaviors in human-like ways with some individuality (cf. Chapter 7.8) conveys the impression of the presence of a mind, which may enhance the agent’s believability more than any idiosyncratic behavior (cf. Section 10.2).

Concerning the interplay between speech (cf. Section 10.1.1) and gestures in discourse, Buisine et al. (2004: 219) listed three ways in which these two modes may cooperate in the behavior of an agent to convey meaning in a presentation context, which they extracted from the taxonomy of types of cooperation between communication modes developed by Martin et al. (2001):

- *Redundancy*. The agent provides verbal information and repeats the same information non-verbally by performing either an iconic gesture or deictic gesture.
- *Complementarity*. Speech and gesture provide not the same but different pieces of information which have to be merged into an interpretation. For example, the agent may talk about an object and indicate its size or shape with an iconic hand gesture.
- *Specialization*. The entire content of a presentation is conveyed using either speech or gesture. The former happens when the agent consists only of a talking head (cf. Chapter 9.6); an example of the latter situation is an agent who gives its presentation through sign language.

Buisine et al. (2004) conducted an empirical study involving three embodied presentation agents to determine whether these three multimodal strategies, when performed by the agents, would be perceived by human observers and what strategy (if any) would make the most impact. They found (pp. 232f.) that while the majority of participants did not consciously notice the different multimodal strategies, both redundant and complementary strategies were more effective with respect to the perceived quality of the explanation provided by an agent than the specialized condition in which the explanation was given only using speech. No difference in ratings was found between complementary and redundant multimodal strategies. Interestingly, only ratings of *male* participants produced this main effect of the study; the female participants did not seem to prefer any of the three strategies. The same kind of interaction between agents’

[111] In fact, style is a broad concept that manifests itself not only in the agent’s manner of gesturing but also in its appearance, choice of language, way of speaking, etc.

multimodal strategy and participants' gender was found in the ratings of trust in the agent (effect for males but not for females). No effects of the multimodal strategy were found for likeability, personality and expressiveness, and recall performance. Overall, the results of the study suggest that complementary and redundant multimodal strategies lead to more favorable perceptions of the quality of a presentation than relying on speech only. As the subjects' performance did not seem to be affected by the multimodal strategy chosen, this study lends support to the hypothesis that embodied agents tend to affect subjective rather than objective variables (van Mulken et al. 1998).

10.1.3 Face

Anatomically, the face is the front part of the head of a human being or an animal, which extends from the forehead to the chin and from one ear to the other. The components of the human face include hair, forehead, eyebrows, eyes, nose, cheeks, mouth, lips, teeth, skin, and chin. These features combine to convey various aspects of a person's appearance, identity, and expression, which distinguish the individual from other people and help others to identify, categorize, and understand that person in social interactions (Donath 2000; Donath 2001; Pantic 2005). The face communicates essential social information about who someone is and how attractive he or she is[112] (through facial features[113]), what is on his or her mind at the moment (through facial expressions), and where his or her attention lies (through direction of eye gaze, cf. Section 10.1.4).

The face provides a unique representation of a person's individual identity. Humans are adept at recognizing and remembering faces. They have the cognitive machinery to identify other people by their face, at a distance, in crowds, from various viewpoints, when the face shows different expressions, and as the face changes with age (Zebrowitz 1997). The close tie between the face and individual identity has become manifest even in language. According to dictionaries, to be faceless is to be "anonymous, without character or identity" (Donath 2001).

The face is not only important for recognizing others as individuals but also for classifying them "at a glance" in terms of social categories including age,[114] sex,[115] race and ethnicity,[116]

---

[112] It is common wisdom that the face plays a central role in judging attractiveness, one's own and that of others (Green et al. 2008). Branham pointed out that "one would be hard-pressed to name one culture that did not in some way encourage its members to alter the appearance of their faces" (Branham 2001) in order to enhance their aesthetic appeal. She quoted Liggett, who observed that "[b]eauty must be pursued at whatever price, because it confers on its possessor profound social influence, power and respect" (Liggett 1974: 46).

[113] While cheek width, eye separation, face height, and jaw width play a central role in attractiveness judgments of female faces, face height and jaw width are important proportions for judging the attractiveness of male faces (Green et al. 2008: 2460).

[114] Progressing age leaves its marks on a person's face. Especially around the forehead, the mouth, and the eyes, the facial skin shows more wrinkles, increased dryness, and reduced elasticity. The eyes and the mouth droop as a result of the lower part of the face growing forward and downward. Therefore, older people may appear less attractive to observers (Nass & Brave 2005: 139).

[115] In comparison to men, female faces are typically wider with a smoother skin, a sharper ridge above the eyes, thinner and higher eyebrows, a larger distance between the eyebrows and eyes, larger pupils, and rounder (tapering) foreheads, jaws, and chins (Nass & Brave 2005: 137). Using these facial cues (as well as additional cues provided by the body), people can reliably identify biological sex very shortly after seeing an unknown person. They also prefer consistency in the facial indicators

(sub-) culture, etc., as well as for making judgments about their personality[117] (Donath 2001). This quick categorization influences one's expectations about and behavior toward other people, and the interpretation of their actions and language. However, such classifications are usually subjective and often based on stereotypes associated with different groups or 'types' of people rather than deeper knowledge about the individual, resulting in significant biases toward the other in the interaction. For example, someone with foreign-looking facial features is often assumed to have a different native language and to be less competent in one's own language than oneself. Likewise, a baby-faced adult will be stereotyped as being trusting, naïve, kind, and weak, in short, child-like in nature (Zebrowitz 1997). Many physiognomic judgments of personality (like this one) are based on people's strong responses to certain visible attributes and emotions (e.g. age, health, anger, etc.), which are overgeneralized to individuals with faces that exhibit some resemblance to such an attribute or emotion (Zebrowitz 1997; Isbister 2006: 10ff.).

In face-to-face interactions, facial displays are synchronized to one's own speech or to the speech of others. They can both replace sequences of words and accompany them (Cassell 2000a: 5). People use facial cues (gaze and expression) from other parties to time and modulate their contributions to an ongoing conversation (Donath 2001). Isbister wrote that "[i]ntentionally displayed facial expressions help convey intentions and relationships to others and fulfill social obligations to have certain feelings at certain times (for example, the persistent smile of the flight attendant)" (Isbister 2006: 152). She also provided some examples of expressions that communicate social intent: friendly or suspicious expressions, dominant or submissive facial reactions, and ongoing facial reactions to shared experiences and stories (p. 153).

Facial displays can also help interlocutors to disambiguate the verbal message when the acoustic signal is degraded, due to the complementary nature of auditory and visual information in speech (Cassell 2000a: 5). The visible articulation of speech, involving observable coordinated movements of the lips, tongue, teeth, and jaw displayed on the face, provides essential information for both normal and hearing-impaired individuals (Massaro 1998). Animated agents with the ability to analyze auditory and visual speech and to produce realistic audio-visual speech using a talking head (cf. Chapter 9.6) can help learners with hearing impairment to improve their speech perception and production skills.

Furthermore, the face is the primary vehicle for the visual communication of an individual's emotional state. The mouth, eyes, and eyebrows are the central elements for portraying emotional messages on the face. A set of basic emotions (anger, disgust, fear, happiness, sadness, surprise, and embarrassment, cf. Chapter 7.15.7) has been identified, which are

---

of sex: faces (and voices) that are clearly male or female are considered more attractive than mixed faces (p. 138).

[116] Skin color is arguably the most obvious indicator of the place of origin of a person's ancestors, but there are other relevant facial cues, such as hair color and the shape of the eyes and the nose (Brown et al. 1998). Concerning the relationship between race and ethnicity, it should be noted that while the two are related concepts, ethnicity is related to the notion of social grouping, whereas race is rooted in the idea of biological classification ([INT 84]).

[117] For example, people with baby-face features (large eyes and pupils, small chin, high eyebrows and forehead, small nose, and full lips and cheeks) are commonly perceived as warmer and more trustworthy but also as more dependent, less responsible, and more submissive and manipulable than people with a more biologically mature face. This is because humans tend to overgeneralize stereotypical attributes of children to adults with child-like features (Isbister 2006: 10).

recognized across cultures. However, people in different cultures do not always show their emotions in the same way. *Display rules* (Ekman & Friesen 1969; Ekman & Friesen 1975) are cultural conventions that "influence how, when, what, and with whom certain expressions should be displayed and suppressed" (Allbeck & Badler 2004: 110). For example, Japanese people tend to conceal negative expressions with a smile when a person of authority is present (Ekman & Friesen 1969). In general, facial expressions are the most visible but also the easiest controllable expressions of all (cf. Chapter 13.4.1). Therefore, having recognized a certain facial expression does not necessarily mean that the individual's emotional state has been recognized accurately as well (Picard 2000: 175). Furthermore, the correct interpretation of facial displays typically requires human observers to take contextual information into account, including both the broad cultural context and the immediate context of the interaction since facial expressions carry multiple meanings and have to be interpreted in context (Fernández-Dols & Carroll 1997). For example, a smile may be happy, sad, nasty, false, forced, etc., depending on the context; it may be directed at the interlocutor, or the person may smile to himself or herself about some private matter.

People behave socially not only toward real faces but also toward mediated faces (Walker et al. 1994; Koda & Maes 1996; Sproull et al. 1996; Nass et al. 1998). Nass and Brave argued that "[f]aces are the most anthropomorphic features that any interface can have because no other part of the body can reveal as much information about the person being observed" (Nass & Brave 2005: 125). Hence, users will also look for and read social information in the faces of embodied agents.

Creating the right expression and impression with an animated face, in the sense of communicating social messages that are contextually appropriate and consistent with the agent's role and public image, is a complex task, given the expressiveness, subtlety, and meaningfulness of faces (Donath 2001). Three aspects of the face convey social information (by themselves and through their interplay) and thus have to be modeled in the design of faces for embodied interactive software agents: structure, dynamics, and decorations (Zebrowitz 1997; Donath 2001; Pantic 2005).

Structural qualities of the face allow people to recognize and socially categorize others, as well as assess their personality. Structural cues of the face include both static and slowly changing aspects (cf. Table 13).

Dynamic qualities of the face enable people to determine others' emotional expressions and direction of attention. They include gaze direction, pupil dilation, blushing, nose wrinkling, and different shapes of the eyes, mouth, and eyebrows that produce smiling, squinting, frowning, pouting, wincing, and other expressions. Finally, eyeglasses, makeup, hair style, jewelry, and other *decorations* added to a face indicate group affiliation, class distinctions, subculture membership, and the person's individual style.

Facial structure, dynamics, and decorations convey messages not only by themselves but also through their interplay (Donath 2001). For example, haircuts can cause a person to be assessed as younger or older. Categorizing someone as female or male based on structural cues influences the cultural interpretation of decorations like very short (or long) hair or use of makeup. Both structural features and decorations can modify the production and interpretation of facial displays of emotion (e.g. a smile on an attractive vs. unattractive face, hiding one's eyes (and emotions) behind dark sunglasses, etc.).

The human brain processes faces (and voices) differently than all other objects (and sounds) (Nass & Brave 2005: 127). This biased processing seems to give humans the instinctive ability to see faces in all kinds of shapes. The observation that simple line drawings and many non-human objects can manifest human faces is the foundation of Topffer's Law (Gombrich 1972),

Table 13. Static and slowly changing structural aspects of the face. Compiled from Donath (2001).

Static	Slowly Changing
• The overall head shape • The size and placement of the eyes and other features	• The lines and texture of the skin • The color and quantity of scalp and facial hair

which goes back to the artist, designer, and amateur psychologist Rodolphe Topffer (1799–1846), who some regard as the father of the modern comic book (Mishra et al. 2001). Topffer's Law states that "[a]ny human face, however poorly and childishly drawn, possesses necessarily, by the mere fact of existing, some perfectly definite expression. In other words, any squiggle which we can interpret as a face will have a distinct individual personality" (Mishra et al. 2001). The most important elements signaling the presence of a face are "a clearly defined head, eyes and eyebrows (or the equivalent) at the top, and a mouth at the bottom" (Nass & Brave 2005: 137).

While humans have a very liberal definition of what a face (or voice) is, do they distinguish between natural human and synthetic (or humanoid, i.e. human-like, but clearly artificial) faces in the same way as they do between human and synthetic voices (cf. Section 10.1.1)? Nass and Brave described an experiment conducted by Gong (2000) to test the effects of consistent vs. inconsistent humanness of face and voice (Nass & Brave 2005: 131ff.). The experiment used four agents representing different combinations of human or humanoid face and voice:

- Human face with human voice;
- Human face with synthesized voice;
- Synthesized face with synthesized voice;
- Synthesized face with human voice.

The results (pp. 133f.) demonstrated that faces are categorized as either human or humanoid by people. Both a human face with a human voice and a synthetic face with a synthetic voice were regarded as significantly more consistent than the mixed face/voice combinations, leading to the disclosure of more information and ratings of the consistent combinations of face and voice as more trustworthy. In contrast, participants felt uncomfortable with the inconsistent face/voice combinations used in the experiment. Female participants tended to be considerably more influenced by (lack of) consistency between face and voice than male participants. The negative inconsistency effects from mismatching categories of face and voice were replicated in a later study by Gong and Nass (2007), suggesting that people treat humanoid entities differently from natural human ones.

These findings should not be interpreted as a recommendation to avoid synthetic faces (and voices) altogether in the design of embodied interactive agents (Nass & Brave 2005: 136f.). After all, many successful and compelling characters have already been created which are clearly not human, such as Bugs Bunny ([INT 50]) and Mickey Mouse ([INT 51]). However, the faces of all of these characters possess the key features of a human face: eyes, eyebrows,[118]

[118] According to Ekman (1979), eyebrow movements serve as emotional and conversational signals in human communication. In conversation, eyebrow movements can emphasize both individual words and sequences of words as they are spoken. However, the exact role of eyebrow movements in communication remains unknown. As a result, developers of embodied conversational agents have different views about when and how often the eyebrows of their agents should move (Pelachaud et al.

and a mouth, embedded in a clearly defined head (see above). In addition, each of these characters has a unique voice, although the voices are not human sounding. The lesson from these characters and the studies above is that what matters is *consistency* between the face and the voice of an agent (Nass & Brave 2005: 137). A human face speaking with a synthetic voice and a human voice attached to a synthetic face both disturb people, whereas consistent pairings of face and voice categories are preferred. From the perspective of voice interfaces, Nass and Brave summarized the need for consistency as follows:

> *[I]nterfaces should present a consistent social 'face' to users: voices that talk, behave, and look like they sound are much more desirable, intelligent, and comfortable interaction partners. (Nass & Brave 2005: 183).*

10.1.4 Gaze

Eye gaze, the direction where an individual is looking with his or her eyes, provides a rich source of social information in face-to-face interactions between humans as well as between humans and embodied interactive agents (Argyle & Cook 1976; Kleinke 1986; Cassell et al. 1994; Torres et al. 1997; Cassell 2000a; Colburn et al. 2000; Poggi et al. 2000; Donath 2001; Knapp & Hall 2005), which human observers can perceive and interpret proficiently. In social interactions, interlocutors process gaze as both output and input (Donath 2001),[119] using gaze patterns to serve various coordinating, regulating and signaling functions.

First, gaze can help to clarify the verbal message by accompanying ambiguous references (e.g. "I am talking to *you*.", "What is *that*?", "Put it *there*.") (Rutter 1984). Using deictic gaze, speakers can point at objects or persons in the current spatial context. Likewise, embodied interactive software agents can use a combination of deictic gaze and gesture to direct the user's attention to particular objects in the shared environment. As an agent interacts with objects and users, its gaze may need to shift back and forth between them (Rickel & Johnson 2000: 97f.). Furthermore, the eyes can indicate properties of objects. For example, narrow eyes may occur during references to small objects, whereas wide eyes can indicate large objects (Poggi & Pelachaud 2002).

Second, both speakers and listeners use gaze to send social signals (Donath 2001; Lee et al. 2002: 638):

Speakers

- Add emphasis to words, phrases, and utterances by glancing at listeners;
- Direct their gaze at the recipients of their verbal messages;
- May look at listeners to demand their attention or with the intention to persuade them;
- May look away from the listener because they are embarrassed or uncomfortable;

1996; Cassell et al. 2001b). Furthermore, it is not clear whether or not the communicative role of eyebrow movements differs from one language to another. Krahmer and Swerts (2004) conducted an experimental study to investigate the role of eyebrow movements for the perception of focus and whether this role is the same across languages (in particular Dutch and Italian). They found that the role of eyebrow movements for the perception of focus is secondary to the one of pitch accents and that the placement and function of eyebrow movements are language-dependent.

[119] People look at interesting persons or things, and the visible direction of their gaze indicates their current focus of interest (Donath 2001).

- Avert gaze when they engage in complex cognitive processing (e.g. remembering information or thinking about what to say next).

Listeners

- Glance at the speaker to show their continuing attention, interest, or involvement (Knapp & Hall 2005);
- Can communicate a high level of concentration, but also indifference or dislike, through averted gaze.

Third, gaze is used to set up a channel in order to obtain information from other parties in the face-to-face interaction (Argyle & Cook 1976). Speakers direct their gaze at listeners during pauses to determine the extent to which the listeners understand and agree with them while listeners continuously observe the speaker's face and the direction of his or her gaze (Donath 2001; Lee et al. 2002: 638).

The fourth communicative function of gaze is to assist in managing conversational flow (Duncan 1974; Goodwin 1981; Argyle 1988; Cassell et al. 1994; Torres et al. 1997; Poggi et al. 2000; Lee et al. 2002: 638; Poggi & Pelachaud 2002):

Speakers

- Avert gaze when beginning a long utterance and look at the listener when they are done;
- Look away from the listener at the beginning of the thematic part of an utterance and look toward the hearer at the beginning of the rhematic part of the utterance (Cassell 2000a: 20f.);
- Establish mutual gaze with the listener during short conversational turns;
- Look at listeners to signal that they have finished speaking and are willing to turn the floor over to someone else;
- May use gaze to suggest the next speaker.

Listeners

- May direct their gaze at the speaker to indicate that they wish to take the floor;
- May look at one of the participants (including the previous speaker) to encourage him or her to speak next.

Fifth, gaze can communicate emotional state (Cassell 2000a: 5), including both 'social' emotions felt toward others (e.g. admiration, anger, or love) and 'individual' emotions not directed at another person (e.g. fear, joy, or sadness) (Poggi & Pelachaud 2002). Examples include staring at someone in anger or casting one's eyes downward when feeling sad. Finally, gaze is an indicator of personality characteristics (Duncan 1974). For example, a shy person may tend to avoid eye contact with an interlocutor, whereas an extrovert individual may actually seek it.

Gaze is combined with speech (cf. Section 10.1.1), gestures (cf. Section 10.1.2), and facial expressions (cf. Section 10.1.3) to negotiate turn-taking, establish social control, reflect levels of intimacy,[120] and indicate understanding and attention (Donath 2001). Gaze can be unilateral or mutual. The latter occurs when two individuals look into each other's eyes at the same time.

[120] Mutual gaze between strangers is briefer and less frequent than between friends. Prolonged mutual gaze can signal intimacy (Donath 2001).

The length of mutual gaze depends on various factors, including age (longer in adults than young children), gender (longer in women than men), the distance between speakers (longer with more widely separated interlocutors), and culture (e.g. 1–2 seconds for adults in Western cultures) (Argyle & Cook 1976; Colburn et al. 2000).

The use of gaze in communication is governed by complex and culturally dependent rules. Direction, duration, and frequency of gaze patterns differ between contact (e.g. Arabic, French, and Latin American) and non-contact (e.g. German, Japanese, and North American) cultures. In general, members of contact cultures "tend to face one another more directly, interact closer to one another, touch one another more, look another in the eye more, and speak in a louder voice" (Allbeck & Badler 2004: 113). Examples of the use of gaze patterns in contact and non-contact cultures include the following (pp. 113f.):

- Swedes hold eye contact less frequently but longer than the English.
- Hispanic women tend to look directly at others while direct eye contact is typically avoided by Asians, Africans, and West Indians.
- People from Japan, Korea, Thailand, and other Asian countries regard staring as impolite and intimidating. In contrast, looking the interlocutor directly in the eye signals confidence in most Western cultures.
- In Saudi Arabia, interlocutors prefer direct eye contact, but the eyes are usually dull and appear disinterested.
- African Americans have an inclination to look at their interlocutors while speaking and look away when listening, whereas Euro-Americans tend to do the opposite (Ting-Toomey 1999: 126).

Some common patterns of the use of eye gaze seem to exist as well. Too much gaze is usually perceived as a sign of anger or threat, whereas too little gaze can signal shyness but also carelessness.

10.1.5 Posture

Posture is the position or arrangement of the body and its extremities. The term also refers to an individual's characteristic way of bearing his or her body. When they are not moving, human beings may stand, sit, kneel, lie, be on all fours, or hang somewhere. Each of these positions can be classified in further detail. Other, non-humanoid embodiments may share some of these positions and also add their own characteristic postures.

In face-to-face communication, posture signals an individual's degree of involvement, liking for the other participants, and his or her status relative to them (Badler & Allbeck 2001). For example, in interactions among (half-) strangers, an interlocutor may lean forward to convey higher involvement, increased liking, and lower status (Allbeck & Badler 2004: 113). Communicators who have established or are trying to build rapport tend to reflect each other's posture (Knapp & Hall 2005).

Posture is important for conveying the intensity of a number of emotional states (Badler & Allbeck 2001). For example, the perception of sadness or anger in an individual is intensified by a drooping or a rigid, tense body posture, respectively. Then, postures also play a role in attributing credibility or social influence power to members of different cultures. People from the United States tend to link displays of relaxed expressions and postures to credibility and positive impressions of others, whereas restrained facial expressions (cf. Section 10.1.3) and rigid postures are characteristic features of influential people in South Korea and Japan (Ting-

Toomey 1999: 127). Some postures may even be regarded as insulting in certain cultures. For example, in the Thai culture, postures involving a foot pointed at the interlocutor are considered offensive because the feet are regarded as the lowest, most revolting part of the body (Allbeck & Badler 2004: 113).

A well-balanced or erect posture is commonly regarded as an integral part of physical attractiveness. Most cultures view erect posture as an indicator of a well-balanced and adaptable personality. Young boys and men are expected to have erect postures, and women tend to prefer and strive for erect postures themselves as well.

Given the diversity of information communicated by something as deceptively simple as body posture, the different positions of an agent's embodiment, as well as the shifts occurring between them in the course of an interaction, deserve close attention in design. Cassell et al. (2001a) found in their study that posture shifts were correlated with shifts in conversational topic and shifts in whose turn it was to talk.

10.1.6 Locomotion

Locomotion is the event of, or capacity for, self-propelled body movement of a living organism or other entity by which it (typically) transfers itself from one location to another. Depending on the specific capabilities of its body and the requirements of the situation, an individual may perform different kinds of locomotion, including crawling, walking, running, climbing, swimming, flying, jumping, etc. Animated embodied agents can use these locomotive behaviors as well, and others not found in nature, such as rolling around on wheels or teleporting themselves to a different location.

For some applications, it may be appropriate to design embodied interactive software agents as stationary elements in an interface because the agent is expected to remain at a particular location in order to provide a reliable place for assistance and information. Examples include embodied interactive agents in the roles of librarians, newsreaders, and receptionists. Yet, the ability of an agent to move from one place in the environment to another becomes critical for other applications that require the agent to:

- Perform its operations at different locations during the course of an interaction;
- Get out of the user's way in order not to block access to devices or information;
- Move toward objects of interest for manipulation or unambiguous reference;
- Move away from objects that are dangerous or no longer in focus;
- Demonstrate how to perform a particular kind of locomotion.

It might seem enough that an agent is capable of locomotion at all. However, designers should also pay attention to the way in which the agent performs actions that change its location because these behaviors can be performed with different degrees of realism (cf. Chapter 9.5), individuality (cf. Chapter 7.8 and Chapter 9.3), and variability (cf. Chapter 7.9). Locomotion is modulated by various individual and situational factors, including:

- The agent's *personality* (e.g. arrogance, confidence, shyness, etc.);
- The agent's current *emotional state* (e.g. bouncing with happiness, weighed down by sadness, shaking with grief, etc.);
- The agent's *physical constitution* (e.g. athletic, overweight, sick, tired, etc.);
- The current *state of the environment* (e.g. distance, obstacles, terrain, etc.);

- *Cultural differences* (e.g. it is acceptable for Latin male friends to walk arm in arm but not for American men. In the Far East, women traditionally follow one or two steps behind men. Women from Western cultures walk with a longer pace and hold their body more upright, which members of other cultures may perceive as aggressive behavior (Allbeck & Badler 2004: 112)).

It appears that while locomotive movements of the body normally are not performed to convey information (unlike gestures or facial expressions), they do reveal something about the personality, emotional state, physical condition, and cultural background of an individual. In addition, locomotion has been used by people on television to create the illusion of narrowing the distance between themselves and the viewers by walking toward the camera, in order to induce a favorable viewer response (Reeves & Nass 1998: 3f.). Agents could use locomotion to create similar effects.

10.2 Believability-Enhancing Behaviors

Apart from behaviors that facilitate communication, the repertoire of embodied agents often includes another class of behaviors which have no direct relation to the ongoing interaction or the current task but are directed at the agent's persona, aiming to make it appear more life-like and believable. Behaviors intended to enhance the agent's persona are commonly performed when the agent is idle, i.e. not interacting with the user or the environment (Johnson et al. 2000). However, they may also be interwoven with the agent's conversational and task-related behaviors.

Some of these behaviors are reflexive in nature: they serve a biological need and are performed automatically, without volition or conscious control, such as breathing, blinking, or wetting the lips (Cassell 2000a: 19). Agents should perform these reflexive behaviors in a similar fashion, with their natural regularity, to strengthen users' impression of interacting with a living being. Other behaviors are trademarks of the agent's personality (especially idiosyncrasies and signature behaviors, cf. Chapter 7.8) and should therefore occur with sufficient frequency and clarity to be recognizable as such in order to enhance the agent's believability.

While idiosyncrasies and signature behaviors can help to effectively convey an agent's unique identity (Hayes-Roth 2004: 456), which make the agent attractive to users, it is dangerous to overestimate and oversell the effects of these behaviors. Cute can quickly become annoying if it is ill-timed and used excessively, without being backed up by sophisticated communicative and task-related competencies (cf. Chapter 7.1). Care has to be taken to keep a balance between task-related and extraneous behaviors in order to avoid unnecessarily adding to the user's cognitive load. Based on the results of their study, Moreno and Flowerday warned that:

> *Despite the fact that students may prefer to learn with visual APAs [(animated pedagogical agents)] and that designers possess the technology to create them, these representations are not neutral and thus should be made with caution. The most straightforward practical implication of our study is that **multimedia programs should carefully limit the amount of social cues presented in software agents rather than add them for reasons of appeal or entertainment**. (Moreno & Flowerday 2006: 204, emphasis added).*

10.3 Relationship-Building Behaviors

People usually do not only talk business when they engage in conversations with one another. In fact, a significant portion of human-human dialogues is small talk about topics that have no overt relationship with the current task but help the interlocutors to get to know each other better (Bickmore & Cassell 2005). This social dialogue is instrumental in deepening the relationship between the interaction partners (cf. Chapter 5.4.5).

Non-verbal behavior plays a crucial role in social dialogue. It is used to express emotions, to accompany and support speech, to engage in rituals, for self-presentation, and to display interpersonal attitude (Argyle 1988). In particular, the decision to approach a person or an embodied agent and the first impression of them after initial contact has been established are influenced by the positive or negative attitude displayed by the person or agent (Bickmore & Cassell 2005). Both people and agents may signal their attitude through their face, body, and voice. Table 14 lists facial, bodily, and vocal cues which, when displayed by an individual, will show others their hostility, friendliness, or indifference.

Hence, embodied interactive agents that are working to build long-term social-emotional relationships with their users should display non-verbal behaviors that communicate a positive interpersonal attitude, including the following types (Bickmore & Cassell 2005):

- *Immediacy*. Close conversational distance, direct body and facial orientation, forward lean, increased and direct gaze, smiling, friendly facial displays, pleasant facial animation, nodding, frequent gesturing and open body postures;
- *Kinesic*. Head tilts, relaxed body, no random movement, open body positions, mirroring the interlocutor's body posture;
- *Vocalic*. More variation in pitch, amplitude, duration and tempo, reinforcing backchannel feedback (*uh-huh*, *mm-hmmm*), greater clarity, expressiveness, fluency, pleasantness, and warmth of speech, smoother turn-taking.

In the remainder of this section, the role of immediacy in relationship building is discussed in more detail (cf. Section 10.3.1). Furthermore, applications of humor in relationship building between humans and agents are explored in Section 10.3.2.

10.3.1 Immediacy

Immediacy has been defined as the extent to which communicative behaviors reduce perceived distance between individuals and enhance physical or psychological closeness in interpersonal communication (Mehrabian 1967; Thweatt & McCroskey 1996). Both verbal and non-verbal communicative behaviors contribute to immediacy. *Non-verbal immediacy* is closeness produced through the non-verbal behaviors listed above (see also the discussion of agents' non-verbal communicative behaviors in Section 10.1). *Verbal immediacy* is created when speakers express themselves linguistically in ways that reveal their liking to their addressees (Mehrabian 1971; Gorham 1988). Verbal behaviors that contribute to immediacy include asking questions, using humor (cf. Section 10.3.2), addressing individuals by their name, initiating conversation, and sharing personal experiences, among others (Woods & Baker 2004). The perceived characteristics of immediate individuals and their impact on their environment are summarized by Andersen:

Table 14. Cues of friendliness, hostility, and indifference (Isbister 2006: 29).

	Friendly	Hostile	Indifferent
Face	Smile; steady but not overly intense eye contact	No smile; intense eye contact	No smile; not much eye contact
Body	Open, relaxed stance; closer in and may lean toward	Tense stance; may be closer in and lean toward	Closed, relaxed stance; may stay further away and lean away
Voice	Warm, energetic	Cold, energetic	Cold, less energetic

> *The more immediate a person is, the more likely he/she is to communicate at close distances, smile, engage in eye contact, use direct body orientations, use overall body movement and gestures, touch others, relax, and be vocally expressive. In other words, we might say that an immediate person is perceived as overtly friendly and warm. (Andersen 1979: 545).*

For many roles, especially those with a supporting character, it is useful for embodied interactive software agents to exhibit verbal and non-verbal behaviors that allow users to perceive them as immediate persons, whose friendliness and warmth makes users' experience of the interaction more enjoyable and satisfactory.

10.3.2 Humor

Relationships between users and agents can also benefit from the use of *humor*, i.e. of stimuli that produce the experience of amusement and tend to elicit the laughter reflex in other individuals; in other words, which have the quality of being 'funny' in the eyes of a particular audience. Humor is a phenomenon of considerable extensiveness and complexity, which depends on culture as well as the individual and has many different manifestations (puns, jokes, sarcasm, slapstick, pranks, etc.) (Ritchie 2001). Humorous stimuli may be expressed verbally and/or non-verbally, using language, (moving) images, sound, actions, smell, and touch, but certain objects or situations can amuse people as well. Someone's *sense of humor* is his or her trait of being able to perceive, appreciate, and express the humorous. The extent to which an individual will find a humorous offering amusing or comic depends on factors including cultural background, period taste, personal preference, maturity and education, and the style and technique of delivering the humor. In general, people regard sense of humor as a highly valued characteristic of themselves and others (Nijholt 2003: 7). They are also likely to be attracted to someone who appears to have a similar sense of humor. Cann et al. (1997) conducted an experiment in which subjects interacted with an unseen stranger who, as they were led to believe, shared many, or very few, attitudes with them. The participants told the stranger a favorite joke, and the stranger responded positively or neutrally to the joke. The results of the experiment indicated that while both attitude similarity and response to the joke contributed to interpersonal attraction, it was the positive response to the joke that had sufficient power to overcome attitude dissimilarity. A dissimilar stranger who responded positively to the joke was rated as more attractive than a stranger who responded neutrally. Cann et al. concluded that humor appreciation is a critical dimension of interpersonal attraction.

Although humor plays an important role in the interactions and relationships which people have with other people, it is a widely underutilized concept in the design of human interactions and relationships with computers. The computer is stereotypically rational, reliable, and serious, so “humorous computer” is an oxymoron, a conjunction of contradictory terms (Nass & Brave 2005: 145). In the minds of many developers and theorists, humor and computers, like entertainment and “serious work,” are separate – and the twain shall never meet. However, humor does have facilitating functions in everyday human communication that go way beyond idle entertainment (Binsted 1995). People use humor “to criticize without alienating, to defuse tension or anxiety, to introduce new ideas, to bond teams, ease relationships, and elicit cooperation” (Barsoux 1993). Non-offensive humor has been shown to facilitate work, improve socialization, bond employees together, create rapport, and boost morale (Clouse & Spurgeon 1995). Given that humans make use of humor to mitigate all kinds of communication problems, and also given that some of these issues, including patronizing or intimidating advice, irritating ignorance, and repetitive apologies, can arise in human-computer interaction as well, it has been suggested that the judicious incorporation of humorous mechanisms into the user interface can make computers appear less alien, less intimidating, and less patronizing, in short, more user-friendly (Binsted 1995; Nijholt 2006: 62). Binsted proposed that the careful use of certain types of humor in conversational interfaces (cf. Chapter 11.3) can help to mitigate some of their shortcomings (limited coverage of the language, insufficient domain knowledge, and irritating conversational behavior): clarification requests become less repetitive, acknowledgement of gaps in knowledge more acceptable, and error messages less condescending when they are delivered with the proper dose of humor. A growing number of researchers are convinced that “[i]f computers are ever going to communicate naturally and effectively with humans, they must be able to use humor” (Binsted et al. 2006: 59).

Morkes et al. (1999) studied the effects of humor in task-oriented human-computer interaction (HCI) and computer-mediated communication (CMC). Participants were told that they were interacting with another person (CMC) or with a computer (HCI) in another room, via a networked computer; however, all of them received pre-programmed comments which differed only in whether they contained humor or not. People receiving task-related humorous comments during the interaction rated the ‘person’ or ‘computer’ they worked with as more likeable and competent, reported greater cooperation, contributed a larger number of jokes, responded in a more sociable (i.e. polite and friendly) manner, and smiled and laughed more. In contrast, the participants interacting with the computer were less sociable, smiled and laughed less, thought that they were less similar to their interaction partner, and spent less time on the task than those interacting with another person through the computer. In general, humor did not distract people from the task at hand, neither in the HCI nor the CMC condition. Consequently, use of humor may enhance the likeability of an interface without affecting users’ performance, contrary to the widespread belief that humor in HCI is a source of distraction, a waste of time, and a cause of slackness. However, the quality of the humor requires attention. Participants in the Morkes et al. study who received *unsuccessful* humor (i.e. jokes seen as unfunny) responded more unsociably than those who were told *successful* (funny) jokes.

Humor does not only alleviate communication problems and facilitate social interaction but may also have positive effects on cognition. Stock and Strapparava noted that humor facilitates getting and keeping people’s attention (due to its potential to provide varied stimulation) and helps them to remember information (because people can link new information to humorous experiences) (Stock & Strapparava 2006: 65). In general, humor can induce emotional states which, in turn, affect cognitive processing (cf. Chapter 6.4 and Chapter 13.2.1). Furthermore, humor acts on people’s beliefs, expectations, and opinions, by challenging and changing their perspective on things. Humorous situations that ridicule established ideas and viewpoints can open people for new ways of thinking; in other words, humor also encourages creativity (Stock & Strapparava 2006: 65). Finally, it is assumed that an interface using humor motivates users to interact with the computer longer and more often and to contribute more to the interaction. The findings of Morkes et al. described above lend some support to this hypothesis, but it should be noted that the positive effects of humor were smaller when users thought that they were interacting with a computer. It would be interesting to see how adding more anthropomorphic attributes (i.e. face, body, voice, and language) to the representation of a human or artificial interlocutor interacts with the use of humor.

The previous paragraphs have shown that speakers make use of humor to achieve important social ends, although it is also true that humor often happens spontaneously rather than as a result of deliberation. Education is another domain in which humor can demonstrate its strengths. Nijholt summarized the positive effects of humor in teaching and learning as follows:

> *Humor contributes to motivation, attention, comprehension and retention of information, and the development of affective feelings toward content. It helps create a more pleasurable learning experience, fosters creative thinking, reduces anxiety, and so on. (Nijholt 2006: 62).*

Based on the conjecture that the positive effects of humor in interactions between learners and human educators can also be realized in the exchanges between learners and computer-based tutors, some researchers have begun to embed mechanisms for the detection, interpretation, generation, and prediction of humor into pedagogical agents that participate in instructional conversations with learners in subject areas such as second language learning (Binsted et al. 2003) and stock markets (Laufer & Tatai 2004). In these applications, humor facilitates learners' interaction with the agent, but it is also useful for motivating learners and to help them to understand the subject matter better. For example, Binsted et al. discussed several types of linguistic humor, i.e. humor based on and transmitted through language, that can be used in second language learning: puns illustrate the relationship between spelling and pronunciation; jokes based on idioms provide contexts that facilitate memorizing these constructions; and scalar humor ("X is so Y that Z") introduces learners to comparison structures in the target language (Bergen & Binsted 2004). Finally, studying and practicing the humorous forms of the target culture can enable learners to gain deeper understanding of and higher proficiency in communicating in that culture.

However, humor in human-computer interaction can do at least as much harm as it can do good. Designers should be very careful with respect to what types of humor they include in an interface. Examples of appropriate and inappropriate humor are summarized in the following (Nass & Brave 2005: 153). If the humor is aggressive, disparaging, irreverent, or sarcastic, it may be perceived as hostile, especially when a female voice delivers it. Jokes with ethnic, racial, or sexual content should be avoided in general because they are likely to be perceived as offensive and may cause legal problems for both content and interaction designers. Some kinds of humor (intellectual, satirical, and word-play) may be too hard to grasp for users and, as a result, increase their cognitive load (which could be problematic when users are in the middle of some task) and make them feel left out of the joke. On the other hand, there is also plenty of sick, toilet, and vulgar humor which may be too crude and distasteful for the users' comfort. Only light and non-provocative humor is 'innocent' enough to be consistently effective. Self-deprecatory humor is also quite safe (although it lowers the perceived status of the speaker), whereas making the user or a third party the target of humor is a dangerous strategy most of the time because of the risk of causing offense (Binsted 1995). Furthermore, it should be noted that jokes of whatever kind come with an "expiration date;" therefore, they should be replaced with new ones from time to time before they become too old (Nass & Brave 2005: 154).

When an embodied agent is designed to use humor, it is necessary to consider the context in which the humor will be delivered, which includes the agent, the audience (mostly, the user), the situation, and the cultural context (Nass & Brave 2005: 154f.). Jokes have to be consistent with the personality (cf. Chapter 13.5) of the individual that tells them. People tend to assume that extrovert rather than introvert persons tell jokes, and someone with a particular personality is commonly associated with a certain type of humor (p. 154). Jokes have to be consistent with the culture of both the teller and the audience because it is not possible to translate jokes in the same way between languages and cultures as other types of messages (p. 155). The effects of some jokes are tied to a particular situation, and it may be hard to recreate the same effects outside that situation. Furthermore, some domains and circumstances are more conducive to or appropriate for (different kinds of) humor than others (p. 155). Sometimes, use of humor can even be counterproductive because it suggests a lack of involvement, professionalism, or sincerity (as the character Q from the James Bond movies once said, “I never joke about my work”). Agents should also steer clear of humor when the user has a reason to be upset, for example when he or she (or the agent) has made a critical mistake, has missed an important deadline, or has lost something (or someone) valuable to him or her. Humor is both positive and arousing, and unfortunately, the arousal part makes a considerably greater impact on the addressee than the (positive) valence. Therefore, the humor coming from an agent tends to intensify whatever positive or negative affective state the user is currently experiencing, making him or her even more upset if the timing is flawed (p. 155). Following Binsted (1995), humor from conversational agents should be reserved for the appropriate contexts, and it should be tailored to the individual user if he or she interacts with the agent on a regular basis.

Humor is a multidisciplinary field of research which traces many of its ideas back to the Greek philosophers Plato and Aristotle (Attardo 1994). Contributions to the field have come chiefly from philosophy, psychology, linguistics, sociology, and literature. Work on computational models of humor, whether as test beds for theories of humor or for use in computer-based systems, has begun only recently and is still in its infancy (for discussions of computational humor research, see Binsted 1995, Binsted 1996, Stock 1996, Ritchie 1998, Ritchie 2001, Mulder & Nijholt 2002, Nass & Braves 2005: 145–155, and the various short papers in Binsted et al. 2006). The major obstacle to building models of humor which enable computers to interpret and generate humor when interacting with people is that the processing of humorous phenomena beyond the simplest types is not possible without substantial knowledge of the world and advanced reasoning capabilities (cf. Chapter 12). Similar to the situation in most of natural language processing and artificial intelligence,

> *[w]orking programs will be possible, for the present, only where the designer can isolate some limited style of humorous phenomenon which is manifested in some manageable medium (e.g. text), and in which the regularities appear to depend on relatively simple knowledge that can be coded up in some tractable fashion, ideally using existing resources (such as lexical databases). (Ritchie 2001: 131).*

Still, the advantages of having workable computational models of humor, for both human-computer interaction and theoretical research, make research on computational humor a worthwhile enterprise. But designers and researchers should keep in mind that they start from a meager foundation and have to face significant challenges (Ritchie 2001: 131).

10.4 High vs. Low Profile

It could be argued that an embodied interactive software agent which is idle for longer periods of time serves no apparent purpose, wasting cognitive resources of the user and computational resources of the system, and therefore might as well step into the background until its services are needed again. This raises the general issue of whether embodied agents should always be visibly and audibly present in an interface or not. On the one hand, an agent maintaining a high profile provides a known place for assistance and information (cf. Chapter 3.1.4), and users, in particular those studying by themselves in web-based learning environments, may draw encouragement and reassurance from knowing that there is always someone available who cares and can offer help. On the other hand, it is also possible that the user will be annoyed or troubled by the feeling of being watched (all the time). The pressure experienced by the user may be intensified if the agent seems to be openly monitoring his or her every move.

Rickenberg and Reeves (2000) performed an experimental study to investigate the effects of different character presentations on user anxiety, task performance, and subjective evaluations of user interfaces. They found that monitoring by animated characters increased subjects' levels of anxiety and negatively impacted their task performance, in particular for users with an external control orientation, i.e. those who thought that their success was controlled by others. However, monitoring characters also increased users' trust in content. Rickenberg and Reeves suggested that beyond their mere presence or absence, the effects of characters in interfaces depend on their actions, messages, and social presentation. They interpreted the relationship between anxiety and increased trustworthiness as indicating that anxiety (a kind of arousal) can also have positive effects. Arousal can determine memory and focus of attention (Lang 1995). Hence, the right level of arousal in an interface can make users pay attention and remember. According to Rickenberg and Reeves, well-designed animated characters in interfaces can have these effects, increasing interest (and a little anxiety) in ways that enhance desirable social responses. Concerning locus of control, they recommended that animated characters should be avoided when designing an interface for users who are confident that they can accomplish tasks on their own (internal locus of control). However, when users lack this confidence (external locus of control), animated characters can make the interaction more robust. Finally, Rickenberg and Reeves noted that locus of control is a relatively stable disposition of people rather than a fleeting reaction to a specific task, so decisions concerning the use of animated characters in interfaces should depend on characteristics of the user, not the task at hand.

10.5 Summary

This chapter was concerned with the behavior of embodied interactive software agents. The discussion focused on social behavior, which takes place in a social context and results from the interaction between and among individuals. Three functions of social behavior were identified (communication, believability enhancement, and relationship building) and discussed in detail.

The communicative behaviors of embodied interactive agents involve the coordinated use of verbal and non-verbal means provided by the agent's body to regulate the conversation process, contribute content to the ongoing conversation, and convey emotional and cognitive states. In Section 10.1, the design of the following communicative behaviors for embodied interactive agents was discussed: voice, gesture, face, gaze, posture, and locomotion.

Section 10.2 focused on behaviors performed by an agent with the explicit intention to make its persona more life-like and believable. These idle behaviors include reflexive actions as well as idiosyncratic and signature behaviors. It was argued that extraneous behaviors should be kept in balance with task-related behaviors and should not be used in an ill-timed or excessive way.

Section 10.3 discussed verbal and non-verbal communicative behaviors that can be used by embodied interactive agents in social dialogues to facilitate building long-term social-emotional relationships with users. The roles of immediacy behaviors and humor in relationship building between agents and users were given closer attention in this discussion.

Finally, Section 10.4 was concerned with the level of presence that an embodied interactive agent should maintain in its environment. Pros and cons of the constant audible and visible presence of agents were discussed, including experimental evidence obtained by Rickenberg and Reeves (2000).

11 Conversation

Agents with an articulate embodiment (cf. Chapter 9) can perform a wide range of behaviors (cf. Chapter 10) that affect the user and the environment. Furthermore, they can be equipped with appropriate sensory and software components to perceive and interpret changes in the environment and actions of the user, which, in turn, affect their own internal states and behavior. The capacity for mutual action, or 'inter-action' (cf. Chapter 5), is critical for agent-user relationships in general and the relationships between learners and pedagogical agents in particular. In fact, experimental evidence (Moreno & Mayer 2000; Moreno et al. 2000; Atkinson 2002) suggests that pedagogical agents equipped with an articulate body and language skills add value to the learner's experience through their ability to participate verbally and non-verbally in instructional conversations with the learner (cf. Chapter 7.1). Conversational agents have to be able to recognize and interpret, generate and express spoken or (type- or hand-) written messages in one or multiple human languages. Therefore, technologies for processing natural language in these ways are reviewed in Section 11.1. However, human-human communication naturally combines language with other modes of communication, including gesture, gaze, facial expressions, posture, and locomotion (cf. Chapter 10.1), as well as tactile (touch), gustatory (taste), olfactory (smell), and other modes. Multimodal interfaces, discussed in Section 11.2, allow people to interact with machines through combinations of language and behavior captured from at least two input modes, to which the machine responds with multimedia output (Oviatt 2003: 287). The review of component technologies in the first two sections provides the background for the ensuing discussion of different types of conversational agents (cf. Chapter 3.2.1), which integrate technologies for processing natural language and other input and output modes to be able to participate in (embodied) conversations with users. After a review of principles of conversation in Section 11.3.1, chatbots are introduced in Section 11.3.2 as the first, rudimentary type of conversational agents that is seeing widespread use, especially on the World Wide Web. Dialogue agents (cf. Section 11.3.3) incorporate more sophisticated language and dialogue processing techniques to enable more flexible and goal-directed human-agent conversations. Finally, embodied conversational agents increase the communication bandwidth of dialogue agents with their ability to process a variety of input modes and to perform meaningful conversational and other actions using their animated face and body (cf. Section 11.3.4). The chapter concludes with a summary in Section 11.4.

11.1 Human Language Technologies

Pedagogical agents with language skills can communicate with learners in a way that is familiar to them from the world outside the computer, namely by engaging them in a natural language instructional dialogue. To keep up their end of a natural language conversation with the learner, pedagogical agents require *human language technologies (HLT)*, i.e. information processing technologies that make use of knowledge about the structure of natural human languages and the world to enable or improve intelligent processing of these languages (Allen 1994; Cole et al. 1997; Atwell 1999; Dale et al. 2000; Jurafsky & Martin 2000; Uszkoreit 2001; Mitkov 2003; Coleman 2005; Jurafsky & Martin 2008). Intelligent processing of natural language invariably requires that a computational model of language (Allen 1994: 1) lie at the heart of a computer system. The system then uses this model to:

- Transform natural language into an internal representation (*understanding*) and/or
- Transform an internal representation into natural language (*generation*).

Conversational agents, which were introduced in Chapter 3.2.1, are a special kind of system based on human language technologies that interprets and responds to users' natural language inputs occurring in the context and course of a spoken or written dialogue, possibly combined with other input and output modes (Sidner 2002; Lester et al. 2004; McTear 2004). The most important human language technologies involved in building conversational agents are listed below and shown in Figure 40. The following sections summarize the state of the art in these technologies.

- *Speech technologies*. Since speech is a central carrier of meaning in human-human face-to-face interactions, conversational agents require at least the ability to identify the words in the user's speech and to deliver their own messages using speech.
  - *Speech recognition* converts spoken words captured through a microphone to a written representation. An understanding of the spoken message is not implied. Speech recognition is reviewed in Section 11.1.1.
  - *Speech synthesis* (cf. Section 11.1.2) converts input consisting of text or some internal representation of information to spoken language output (*text-to-speech* vs. *concept-to-speech*).
- *Language technologies.* To make sense of the user's contributions and to be able to respond with meaningful contributions of their own, conversational agents require *natural language processing (NLP)* technologies, which extract information from and convey information using natural language.
  - *Natural language understanding (NLU)* translates a natural language input into a formal representation of its meaning, which can be used for further processing. NLU is reviewed in Section 11.1.3.
  - *Natural language generation (NLG)* generates natural language text from a formal representation of content. This technology is discussed in Section 11.1.4.
- *Knowledge management*. Like software agents in general, conversational agents add value to users' experience (cf. Chapter 7.1) by possessing and communicating some form of expertise which users perceive and make use of (cf. Chapter 7.2). Expertise presupposes the ability to manage knowledge about domain concepts and tasks, which includes its acquisition, representation, and use to solve problems that are the focus of or occur during the interaction with the user. Expertise is the topic of Chapter 12.
- *Dialogue management*. The ability to sustain an effective, goal-directed conversation with the user is critical for a conversational agent. The dialogue manager carefully orchestrates the agent-user exchange across multiple turns toward a successful and satisfactory conclusion. Dialogue management is discussed in Section 11.3.3.2.

Traditionally, conversational systems have only used a single medium (text or audio) and a single communication channel (typing or speaking). But the environments of conversational agents are becoming more complex, requiring agents to handle both multiple media and multiple modes of input and output. In these environments, conversational agents also need:

- *Multimedia technologies*. State-of-the-art multimedia environments, like the Virtual Linguistics Campus (cf. Chapter 2.4), provide content that integrates multiple media,

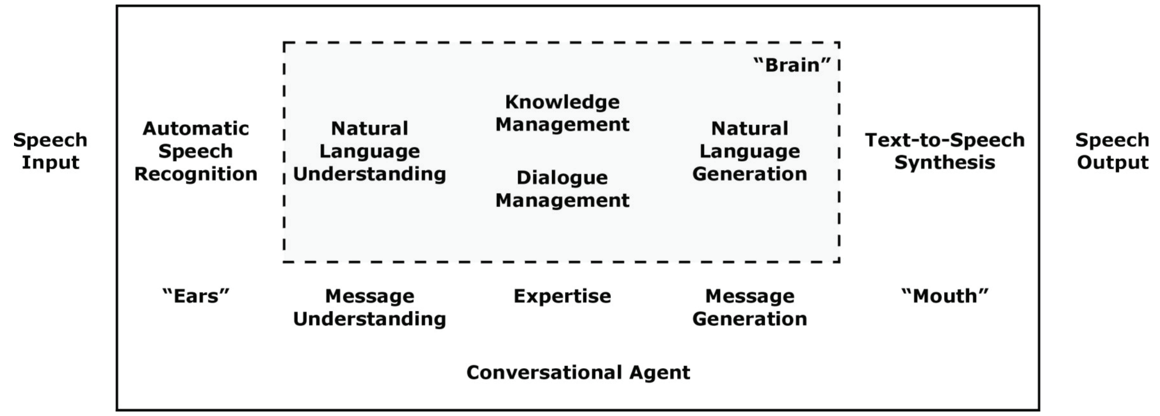

Figure 40. Component technologies of a conversational agent.

including text, graphics, video, animation, and sound, in various ways (cf. Chapter 2.3.7). Agents situated in these environments may have to deal with multimedia content as input, output, or source of knowledge. Hence, they should be able to interpret, manipulate, generate, and refer to multimedia content in the context of their ongoing interaction with the user.

- *Multimodality technologies*. The ability of conversational agents to integrate and coordinate several combined input modes (speech, pen, mouse, facial expressions, gaze, gesture, presence, posture, etc.) and multimedia outputs (combinations of images, text, speech, video, animations, etc.) can increase the bandwidth of human-agent communications substantially and allows agents to compensate for shortcomings of the component technologies involved. Embodied conversational agents (cf. Section 11.3.4) with the appropriate multimodal interface technologies can both interpret and perform multimodal communicative actions that combine spoken natural language with the non-verbal behaviors discussed in Chapter 10.1.

11.1.1 Speech Recognition

Automatic speech recognition (ASR) is the computer-based process of mapping from an acoustic speech signal, which has been obtained by means of a microphone or a telephone, to a string of words in a particular natural language (Rabiner & Juang 1993; Rudnicky et al. 1994; Zue et al. 1997; Holmes & Holmes 2001; Huang et al. 2001; Lamel & Gauvain 2003; Coleman 2005). People can recognize (and understand) the speech of many different individuals across a wide variety of situations with relative ease; hence, it is not surprising that ASR research and development try to achieve this ideal by building machines which can:

- Recognize the speech of any individual without prior training on a voice;
- Recognize any word of one or several languages;
- Handle users' speaking at any pace;
- Distinguish between commands, dictation, and non-machine-directed speech;
- Perform robustly in any situation, especially under unfavorable conditions (e.g. background noise, low signal quality, etc.).

Speech recognition systems that possess these capabilities to the same degree as an average human being have not been built yet. The same task that people perform so well with different speakers in different contexts every day is much more difficult to handle for the average computer, due to the inherent complexity of the speech recognition task, which arises from the nature of the speech signal as an analog, continuous sound wave with high levels of redundancy and noise, and from various task-related issues:

- *Segmentation* is the problem of extracting speech sounds, syllables, and words from the speech signal. Normal fluent speech only rarely marks the boundaries between speech units unambiguously.
- *Variability* in the speech signal means that the pronunciation of speech units is not constant but highly dependent on (Zue et al. 1997: 4ff.):
    - The acoustic/phonetic context[121] of speech sounds (*phonetic variabilities*);
    - The environment (background noise, other speakers' talking, echoes in the room, etc.) and the communication channel (problems with the microphone, poor transmission quality, etc.) (*acoustic variabilities*);
    - The individual speaker, including such aspects as speaking rate, stress, pitch, loudness, intonation, clarity of pronunciation, and disfluencies (e.g. false starts, gap fillers, throat clearing) (*within-speaker variabilities*);
    - Differences between individual speakers and between groups of speakers, including sex, age, dialect, social background, size and shape of the vocal tract (*across-speaker variabilities*).
- *Task-related issues* include people talking over system prompts (*barge-in effects*), their use of words outside the system's vocabulary, the need to separate task-related speech from irrelevant input, the difficulty to distinguish between commands and data in the input, and the user's fleeting awareness and resulting negligent correction of recognition errors.

11.1.1.1 Design Parameters

A number of design parameters can be controlled to make the speech recognition task easier for machines (Kay et al. 1994: 116f.; Holmes & Holmes 2001: 235ff.). One parameter concerns the user's *way of speaking* to the speech recognition system. Speakers can help the system to identify their words by inserting a significant pause after each word. This is called *isolated word* or *discrete speech*. However, discrete speech is unnatural for humans and therefore difficult to produce and annoying for listeners. Furthermore, discrete speech input is also slow. While discrete speech is sufficient for simple tasks like voice command, more sophisticated applications, such as dictation and spoken dialogue, require *continuous speech* recognition, which handles speech as a continuous stream of words without artificial pauses.

Another major parameter that distinguishes speech recognition systems concerns the range of speakers that can be handled, in other words the degree of *speaker dependence*:

- *Speaker-dependent* systems are tuned to the speech patterns of an individual speaker and can only recognize his or her speech with acceptable accuracy.

[121] The articulation of adjacent speech sounds (within a word or across word boundaries) is not strictly sequential but overlaps to some extent because the configurations of the tongue and other articulators are influenced by neighboring speech sounds (Holmes & Holmes 2001: 4). For example, the initial consonant /k/ in *key* is pronounced differently than in *caw*, influenced by the following front vs. back vowel. This pervasive phenomenon is referred to as *coarticulation* (Ladefoged 2001).

- *Speaker-adaptive* systems adapt to the speech of individual users through additional training and thus can increase their recognition accuracy.
- *Speaker-independent* systems accept spoken input from many different people without prior training for each speaker.

Speaker-dependent speech recognizers are easier to develop and perform better than speaker-independent systems, but they are also much less flexible. Speaker-adaptive programs seek to combine the advantages of both extremes; however, the user still has to familiarize the system with his or her voice to improve performance.

A third parameter for classifying speech recognition systems is *vocabulary size*. The number of words a speech recognizer needs to know depends on its task, ranging from two to three words for simple voice command applications to hundreds of thousands of words for current dictation systems. Restricting the vocabulary can make the recognition task easier because there are less potential word candidates to choose from. Controlling for similar-sounding words can also help to improve recognition accuracy (Holmes & Holmes 2001: 236).

Finally, it is possible to control the task and the language that the speech recognizer has to deal with. A task-related measure of the complexity of a recognition problem is its *perplexity*, i.e. the average number of word alternatives the speech recognizer has to consider at any one point in the recognition process (Kay et al. 1994: 119; Holmes & Holmes 2001: 197f.). A lower perplexity means that searching for the correct word is easier and hence more efficient and accurate (Holmes & Holmes 2001: 198).

Restricting the application *domain* of a speech recognition system can help to keep perplexity low and the vocabulary small because it constrains the language (words and constructions) that the system can reasonably expect from users. In fact, early success stories of speech recognition were systems for generating standardized legal or medical reports.

Conversational agents should accept continuous input from a wide range of speakers, without requiring (extensive) prior training for each speaker; in other words, they should be speaker-independent or at least speaker-adaptive. The vocabulary size and perplexity depend on the agent's domain; for most conversational tasks, they will be moderate to high.

11.1.1.2 Speech Recognition Techniques

Most state-of-the-art speech recognition systems adopt a *pattern recognition* approach (cf. Figure 41). Pattern-matching speech recognizers use the output of an initial *signal processing* stage (typically a sequence of *feature vectors*) to *search* a database of reference patterns for the best match, using language-specific constraints on the occurrence of words in different contexts provided by a *language model* to decide what sequence of words has most likely generated the feature vectors observed in the input.[122] The approach is *data-driven*: both the reference patterns and the language models are derived by automated training procedures from representative sets of spoken and written language data.

[122] The various sources of uncertainty associated with the input speech in combination with their own inferior comprehension capabilities make it very difficult for ASR systems to be as confident about what the user said as another human might be.

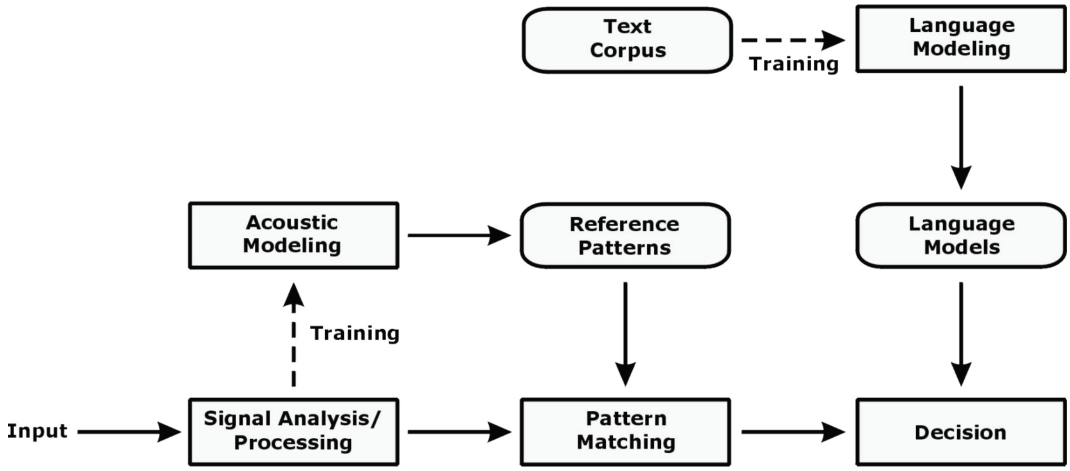

Figure 41. A block diagram of pattern-matching speech recognition. Adapted from Brewer (1996).

Reference patterns have been represented as templates,[123] stochastic models, connectionist networks, or hybrids of stochastic and connectionist models (Kay et al. 1994: 124–138). Of these, stochastic models have enjoyed the greatest success so far, being used in virtually every modern speech recognizer. Stochastic-based approaches define a probabilistic model for each word or other unit to be recognized, which can produce sequences of feature vectors that represent different examples of uttering the word or unit. Given an observed sequence of feature vectors in the input speech, the recognition task then involves computing for each word model the probability that it could have generated that sequence, based on the known properties of the model, and picking the model/word with the highest probability (Holmes & Holmes 2001: 128). The most commonly used stochastic formalism is the *hidden Markov model (HMM)* (Rabiner & Juang 1993: Chapter 6; Kay et al. 1994: 131–136; Holmes & Holmes 2001: Chapter 9; Huang et al. 2001: Chapter 8; Jurafsky & Martin 2008: Chapter 9), which represents models of speech patterns as a network of states connected by transitions, in which each state is associated with probabilities that determine the next state (*transition probabilities*) and what feature vector to emit in the current state (*emission probabilities*). Both transitions and emissions are picked randomly as the HMM progresses from one state to another over a series of discrete time steps, governed by the current distribution of the transition and emission probabilities, respectively (Tebelskis 1995: 16). The performance of hidden Markov models depends on the quality of both sets of probabilities. The values of these parameters are automatically generated in a training process from a set of speech data obtained from a variety of human speakers.

Connectionist approaches apply *artificial neural networks (ANN)* (Rabiner & Juang 1993: 54–65; Tebelskis 1995: 51–71; Holmes & Holmes 2001: Chapter 13; Huang et al. 2001: 457ff.) to the recognition of input speech patterns. ANNs are networks consisting of a large number of highly interconnected simple processing units capable of operating together in parallel to

---

[123] Template-based speech recognition systems use stored sequences of feature-encoded speech frames (so-called *templates*) for each word in the vocabulary (Kay et al. 1994: 126ff.; Fellbaum 1997; Holmes & Holmes 2001: Chapter 8). Templates are only used in older systems with limited capabilities (Rudnicky et al. 1994) because of several serious shortcomings, including their speaker-dependent nature, their inability to handle continuous speech, and their unsuitability for large vocabularies.

achieve complex behaviors or tasks. The connections between units are associated with weights that determine the behavior of the network. These weights can be adapted by learning the underlying patterns of training data (Cawsey 1998: 156f.; Holmes & Holmes 2001: 214). Connectionist models have been successfully used to classify short, isolated speech units (especially phonemes and words) (Bourlard & Morgan 1997: 358), but they have significant difficulties with modeling the speech signal over time. Thus, ANNs do not scale up easily to the recognition of continuous speech in phrases or sentences.

Hybrid approaches to speech recognition combine stochastic models, in particular hidden Markov models (HMM), and connectionist networks (Tebelskis 1995: 57ff.; Holmes & Holmes 2001: 217f.; Huang et al. 2001: 457ff.), in a way that exploits the strengths of each approach. HMMs are used to model the speech signal over time while neural networks are used for signal classification. Hybrid ASR systems have proven superior to classical HMM-based approaches in a number of cases.

The language models used in most current speech recognition systems are statistical models consisting of probabilities of words to occur in contexts of one or two preceding words (*bigram* or *trigram* models) (Holmes & Holmes 2001: 196f.). Using a trigram model, the overall likelihood $P(W)$ of a recognized word sequence W is obtained by multiplying the probabilities $P(w_x|w_{x-2}\ w_{x-1})$[124] of the trigrams in that sequence (p. 197). The string with the highest computed probability wins. Statistical language models are derived from the frequencies of occurrence of bigrams or trigrams in a representative and typically very large machine-readable text collection (*corpus*), which are estimated and used to compute the probabilities of the word sequences during a training process (pp. 185, 197 + 199). Since no corpus covers all possible bigrams or trigrams of the language, many probabilities will be zero, which means that word sequences containing the corresponding bigrams or trigrams cannot be recognized. Hence, *smoothing* techniques are used to raise the zero or low non-zero probabilities and to lower the high probabilities, yielding more uniform ('smoother') probability distributions for the language model (p. 199).

11.1.1.3 State of the Art

Automatic speech recognition is among the fastest growing and commercially most promising human language technologies. It is being used in an increasing number of applications, including command-and-control, dictation, data entry and retrieval, toys and games, and telecommunications (Holmes & Holmes 2001: 239–242). In education, speech recognition technology has a number of useful applications in the field of (foreign-) language learning. It offers learners interactive speaking practice, assessing both *what* they are saying and *how* they are pronouncing it. Promising applications of speech recognition for language learning include pronunciation training (Eskenazi 1999), conversational interaction (Ehsani & Knodt 1998), and reading tuition (Mostow & Aist 2001; Mostow & Beck 2007). In a broader context, ASR could also assist learners by automatically transcribing lectures and speech data when learners find this difficult for cognitive, physical, or sensory reasons; help them to manage and search collections of digital multimedia content; and caption speech for hearing-impaired learners or when speech is not available, convenient, or audible (Wald 2006). Another emerging application of speech recognition is in (embodied) conversational agents (e.g. Ball et al. 1997;

[124] That is the probability of seeing word w_x given the sequence of the two previous words w_{x-2} and w_{x-1} (Holmes & Holmes 2001: 197).

Cassell et al. 2000b; Gratch et al. 2002; Cole et al. 2003; Johnson et al. 2004b), where it facilitates symmetric multimodal human-agent interaction (cf. Chapter 11.3), giving agents the ability to listen to users in addition to talking to them.

The best modern speech recognizers are capable of *large vocabulary continuous speech recognition (LVCSR)* with near-100% word-recognition rates under good conditions (quiet environment; appropriately positioned, high-quality microphone; same speaker), without the need for (extensive) speaker-specific training. However, some limitations of speech recognition technology still need to be removed:

- ASR systems must be able to maintain good performance when input conditions are less than optimal. Changes in the acoustic environment, transmission channel, speaker identity, speaking style, etc. still often lead to sudden and significant degradation in performance.
- Speech recognition technology has to be integrated with other human language technologies because future speech applications, including conversational agents, will require capabilities of understanding beyond the recognition of the spoken words.
- Interactive applications require the ability to handle *spontaneous speech* (rather than read speech), so the speech recognizer must be able to deal with disfluencies like gap fillers, hesitations, self-corrections, coughs, tongue and lip clicks, etc. (cf. Chapter 11.3.3.1).
- Speech recognition uses long-established techniques for acoustic and language modeling. However, it appears that hidden Markov models and bigram/trigram statistics are reaching their limits, so substantial innovations beyond these methods are needed.

11.1.2 Speech Synthesis

Speech synthesis is the process of generating spoken utterances from symbolic input. Depending on the nature of the input, two major types of speech synthesis can be identified. *Text-to-speech (TTS)* synthesis systems (Dutoit 1997; Sproat 1998; Holmes & Holmes 2001; Huang et al. 2001; Dutoit & Stylianou 2003; Tatham & Morton 2005) interpret text in orthographic form (from a computer file or from printed text) and convert it to speech. *Concept-to-speech (CTS)* synthesis starts from some representation of the ideas or concepts to be communicated and expresses them in spoken language (cf. Section 11.1.2.3).

The gold standard for speech synthesis is a system that can deliver arbitrary messages in real time and in a synthetic voice which is at least acceptable to human listeners. The quality of synthetic speech can be assessed along the dimensions of intelligibility, naturalness, and expressiveness, which were discussed in Chapter 10.1.1.

11.1.2.1 The TTS Problem

The majority of current speech synthesizers are text-to-speech systems. To produce a highly intelligible and natural-sounding rendering of a text, a TTS system must deal with both aspects of the input text and desired qualities of the spoken output. Problems with text processing arise from the fact that input texts are neither unambiguous nor give away all necessary information (Sproat et al. 1999: 17ff. + 22ff.; Holmes & Holmes 2001: 97ff.). White spaces, punctuation, and formatting typically do not consistently indicate the boundaries of words, sentences, paragraphs, and other structures. Abbreviations, numbers, symbols, dates, times, and combinations of these have to be expanded into sets of pronounceable words. The pronunciation of proper names, foreign-language terms, scientific and technical terms, word creations, and homographs[125] can only be guessed in many cases or depends on the context. Stress assignment at the word and the sentence level is determined by a variety of linguistic factors, including lexical category, the semantic relations between words, the pragmatic function of a word, and the properties of the discourse (Sproat et al. 1999: 22f.). Longer sentences need to be split into smaller intonational units (*prosodic phrases*), and the sentence type has to be identified. Finally, to convey the author's message properly, the TTS system would have to be able to *understand* the meaning of the text prior to reading it to the user (cf. Section 11.1.3).

Speech generation is difficult because machines lack both the fully-developed vocal tract and the spoken language processing capabilities of human beings. First, synthetic speech has to incorporate the contextual variation of speech sounds arising from coarticulation. While a considerable problem for speech recognition systems (cf. Section 11.1.1), coarticulation effects are indispensable for achieving intelligible synthetic speech (see Hess (1996: 3f.) for a detailed discussion). Speech generation does not work by simply stringing phonemes[126] together (Portele 1997: 59) but requires further manipulation of the speech signal to account for the articulatory transitions between speech sounds.

When humans speak a message, it does not only matter *what* they say but also *how* they say it. In addition to information about the identities of the speech segments, intelligible and natural speech also contains information about *prosodic* features, including (Holmes & Holmes 2001: 6f.):

- *Pitch*. The output speech should exhibit patterns of pitch variation that distinguish words in tone languages (e.g. Chinese) and indicate sentence type, prominence of words, speaker mood, etc.
- *Intensity*. Changes in perceived loudness create the impression of variation of vocal effort by the synthetic talker, help to identify stressed syllables, and indicate boundaries of intonational phrases.
- *Timing*. The speech synthesizer has to determine the duration of each speech segment such that the output speech mimics the temporal structure of human utterances.

The prosodic structure of an utterance should further indicate the overall *voice quality* (e.g. breathiness, creakiness, softness, and loudness) and *voice identity* (i.e. the unique vocal characteristics of the speaker).

11.1.2.2 The TTS Process

The generation of synthetic speech from text is typically implemented as a process consisting of two major steps (cf. Figure 42). The *text analysis* stage determines the underlying linguistic structure of the input text, which is used as a basis for speech generation. It identifies the boundaries of paragraphs, sentences, and words in the input text; expands non-word items into sequences of pronounceable words; determines word pronunciations using a pronunciation

[125] *Homographs* are words with the same graphemic shape but different pronunciations and meanings. Examples include *read* (present tense vs. past tense), *bass* (musical range or instrument vs. fish), and *bow* (weapon vs. to bend one's knee or body).

[126] The *phoneme* is the smallest unit in the sound system of a language that, when replaced with another unit of the same kind, causes a change in meaning, as in the minimal pair *pin* vs. *bin*.

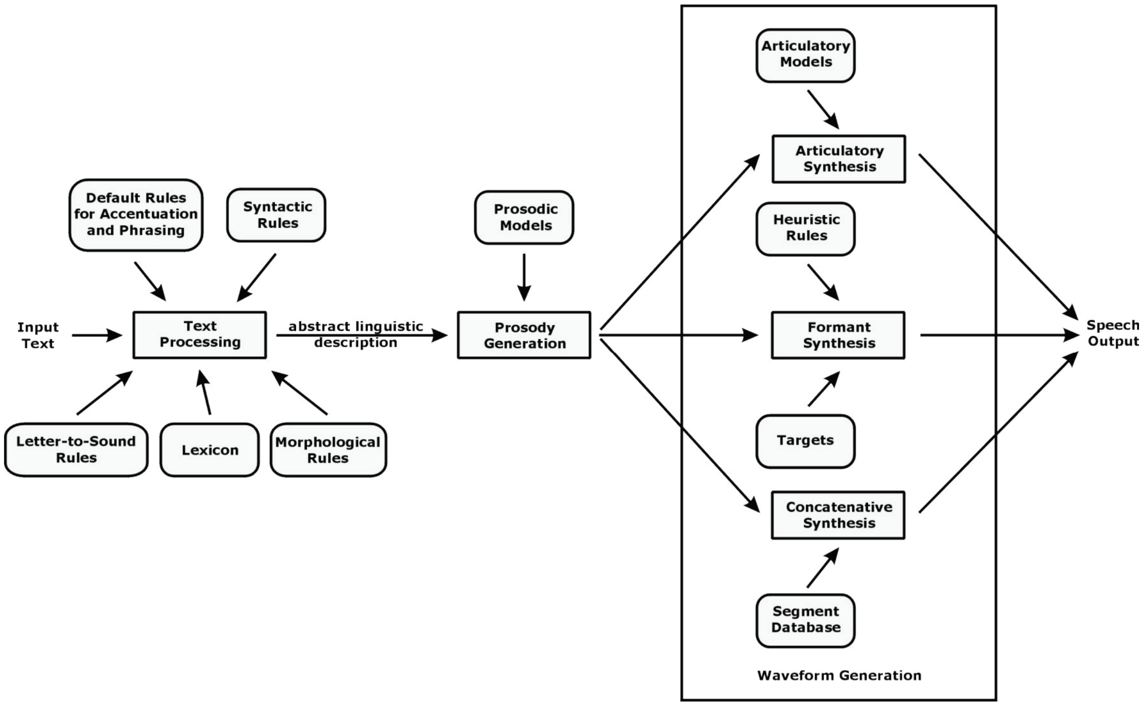

Figure 42. A block diagram of the text-to-speech synthesis process. While the preprocessing of the input text is fairly similar across TTS systems, there are three major approaches to speech waveform generation: articulatory, formant, and concatenative synthesis.

lexicon, *letter-to-sound rules*,[127] and morphological analysis (cf. Section 11.1.3.3); analyzes the syntactic structure of the input text to resolve pronunciation ambiguities, identify the sentence type, and locate accents and prosodic phrases; and, finally, assigns word-level (lexical) and sentence-level stress to the relevant parts of the words and sentences in the input (Holmes & Holmes 2001: 97–102). The final result is a string of discrete sound symbols annotated with information that may influence the way the text is to be spoken.

From this representation, the *speech generation* component produces the output synthetic waveform. First, information in the abstract linguistic description about lexical stress, sentence-level stress, and phrasing is used to generate the *prosody* for the synthesized speech. Speech synthesis systems use a set of rules to model various acoustic correlates of prosody in the synthetic speech signal, including the timing pattern, the overall intensity level, and the fundamental frequency (pitch) contour (Holmes & Holmes 2001: 102–106).

Second, the speech waveform is synthesized by selecting the appropriate synthesis units and then using them together with the prosodic information to produce the spoken output. Speech synthesizers have taken three major approaches to speech waveform generation (Dutoit 1997: Chapter 7; Holmes & Holmes 2001: Chapters 5 + 6; Huang et al. 2001: Chapter 16; Dutoit & Stylianou 2003: 333–336):

- Use of an *articulatory model* (not relevant for practical systems);[128]

[127] Letter-to-sound rules specify how a sequence of letters is to be mapped to a sequence of sound symbols, either abstract linguistic units (phonemes) or acoustic-phonetic units (phones, diphones, or demi-syllables).

[128] For a discussion of articulatory speech synthesis, see Kröger (1996) and Huang et al. (2001: 803f.). Arguments against using articulatory synthesis are provided by Holmes and Holmes (2001: 81f.).

- Synthesis by *rule* (formant synthesis);
- *Concatenation* of speech segments (concatenative synthesis).

Formant synthesis of speech involves modeling the main acoustic features of the speech signal (d'Alessandro & Liénard 1997: 173). It is based on the *source-filter model* of speech production, where a signal generator (the *source*) provides the input signal to a parameterized *filter* that models the vocal tract. Both source and filter are controlled by a large set of heuristic rules (p. 173).

The most important filter parameters are the resonances, or *formants*, which are specified in terms of their frequencies, bandwidths, and amplitudes. Further parameters are necessary to realize voiceless and nasal sounds (Zboril 1995; Holmes & Holmes 2001: 28–31; Huang et al. 2001: 799). Generating the speech waveform involves first specifying a series of acoustic target values to be achieved during the phonetic units making up the utterance and then applying the rules to interpolate between these targets in order to produce an output speech signal that emulates the dynamic aspects of the human voice (Rudnicky et al. 1994: 55; Isard 2001: 793).

Since formant synthesis is based on rules and parameters, it is quite flexible when it comes to realizing different voices or speaking styles (e.g. whispering) (Huang et al. 2001: 803). However, the downside is that the rules for controlling the synthesizer have to be hand-coded and iteratively refined by human experts (Sproat et al. 1999: 27), and a large number of them are required. Furthermore, there is the general problem that no model or rule system has been able to achieve the quality of a human voice so far (Hess 1996: 2; Holmes & Holmes 2001: 107).

Concatenative synthesis techniques join together chunks of stored natural speech available either as waveforms or in some coded format and apply signal processing techniques to smooth the joins between the segments and to modify the prosody of the concatenated sequence of elements in order to match a desired prosodic scheme (Rudnicky et al. 1994: 55; d'Alessandro & Liénard 1997: 174; Jurafsky & Martin 2000: 275f.; Holmes & Holmes 2001: 283; Huang et al. 2001: 818). The units used for concatenation vary in size but are typically larger than isolated speech sounds (phones) to capture the transitions between adjacent speech sounds. Many systems have used diphones, which extend from the invariable part of a phone to the invariable part of the next (Jurafsky & Martin 2000: 274f.; Holmes & Holmes 2001: 71). Others draw on demi-syllables, syllables, words, phrases, or an inventory of segments of different size (Portele 1997; Jurafsky & Martin 2000: 276f.; Huang et al. 2001: 805–809).

A concatenative synthesizer splices only slightly modified chunks obtained from recordings of a single human speaker; hence, much of the quality of natural speech can be preserved in the output (Hess 1996: 2). Not surprisingly, most of today's commercially available speech synthesis systems are concatenative. However, concatenative synthesizers are less flexible than rule-based systems since they are bound to the characteristics of a specific human voice (p. 2). To generate a variety of voices and speaking styles, additional data from the same or from other speakers may be required. In addition, it takes much time and effort to build the database of speech segments for a voice (Hess 1996: 2; Holmes & Holmes 2001: 72f.).

11.1.2.3 Concept-to-Speech Synthesis

Concept-to-speech (CTS) synthesis (Alter et al. 1997; Taylor 2000; Holmes & Holmes 2001: 107f.; Huang et al. 2001: 899ff.; Pan 2002; Tatham & Morton 2005: 321f.) differs from text-to-speech synthesis in that it does not start from unrestricted text input but from a controlled formal representation of ideas or concepts to be expressed in speech (Holmes & Holmes 2001:

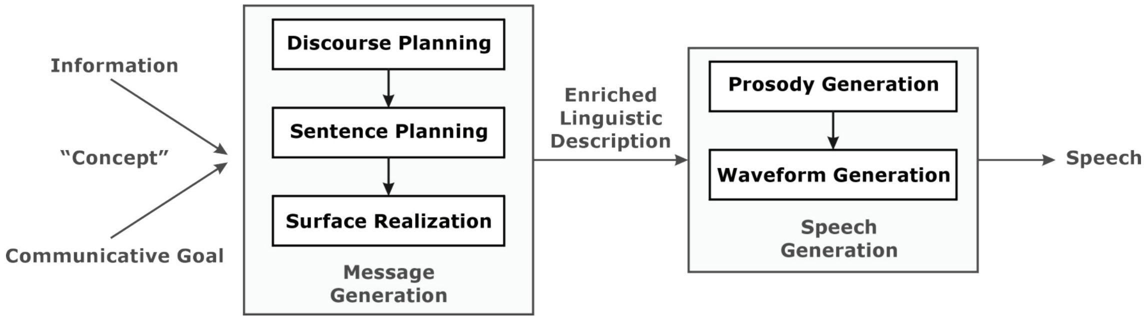

Figure 43. A block diagram of the concept-to-speech process. Adapted from Pan (2002: 7, Figure 1.2). Message generation involves the same sequence of steps as a generic natural language generation system (cf. Figure 45). The stages of speech generation are those found in a text-to-speech system.

107). Usually, such a concept consists of information and a communicative goal, i.e. a specification of both what should be communicated and what this communication is intended to achieve (cf. Section 11.1.4).

Like text-to-speech synthesis, the process of converting concepts into spoken utterances is a two-stage process, which is shown in Figure 43. The first step, *message generation*, involves the production of an enriched linguistic description of the message to be communicated, which specifies in detail what (content selection) is to be said and how (discourse planning, sentence planning, and surface realization). The second step, *speech generation*, synthesizes an output speech signal from that linguistic description, subdivided into prosody generation and speech waveform generation, just like the corresponding stage in a TTS system (Pan 2002: 3–7).

The traditional approach to concept-to-speech synthesis is to have a natural language generation (NLG) component (cf. Section 11.1.4) produce a text from some non-linguistic input, which is then sent to a text-to-speech synthesis system to generate the spoken message (Taylor 2000; Pan 2002: 7). While this approach is simple and convenient, it discards valuable linguistic information after text generation, which has to be recomputed by the TTS system from the text (Zue 1997). As a result, with this approach, nothing is gained for the speech generation stage. The only difference is that the input text is dynamically generated rather than provided directly as input. In contrast, enriching the intermediate description with syntactic, semantic, and discourse information from message generation could facilitate the task of producing output speech with appropriate prosody (Huang et al. 2001: 899ff.; Pan 2002: 9).

11.1.2.4 State of the Art

> *We have reached the point where we can deliver systems that are quite intelligible, so speech system design goals are now shifting to make voice systems sound more like people and less like machines. (Andy Aaron, quoted in Korzeniowski 2008).*

Speech synthesis technology enables interactive systems, including conversational agents, to respond to the user with dynamically generated messages instead of canned (pre-recorded) speech. There are various applications that benefit from the flexibility offered by synthetic speech in comparison to pre-recorded speech, including reading machines and synthetic voices for the disabled; flexible message generation for software systems, devices, and vehicles; speaking toys and game characters; and telephone-based services (Holmes & Holmes 2001: 233ff.). In (foreign-) language learning, speech synthesis could be used for modeling pronunciation, illustrating intonation, and spoken conversation. Further opportunities exist in literacy tuition (Mostow & Aist 2001; Mostow & Beck 2007) and speech therapy (Massaro et al. 2000; Massaro 2005, cf. Chapter 9.6 and Chapter 10.1.3). Since educational applications commonly require high-quality output, recorded human speech is still widely preferred for practical systems. However, as soon as systems have to participate in open-ended spoken interactions with learners, flexible speech generation mechanisms become a necessity. This is certainly true of conversational pedagogical agents (cf. Chapter 10.1.1).

Most state-of-the-art synthesizers generate speech using concatenative techniques. The best systems can produce output of high intelligibility and acceptable naturalness for short, simple utterances (Holmes & Holmes 2001: 107). Depending on the system, the spoken messages can be delivered in different voices and speaking styles. However, even the most sophisticated synthesizers currently available will not pass for a natural (human) speaker for too long because their voices still lack in naturalness and expressiveness, which becomes quite apparent when messages are longer (p. 107). In fact, the perceived quality of synthetic speech can decline rapidly for longer utterances (p. 232), due to its artificial- and neutral-sounding prosody. Wrong pauses, incorrect stress placement, ill-timed pitch changes, occasional mispronunciations of known words, and other lapses all suggest to human listeners that the machine does not understand the messages it delivers, unlike humans (Nass & Brave 2005: 134). Hence, speech synthesis technology is currently most successful in domains in which the utterances are short and have simple prosodic patterns, or in applications that involve more complex utterances but tolerate lower speech quality (Holmes & Holmes 2001: 232f.).

The intelligibility of synthetic speech at the segmental level is largely considered a solved problem: state-of-the-art speech synthesizers can render the speech segments making up the words of the output with sufficient quality for humans to correctly recognize those words (Tatham & Morton 2005: 1). However, improving the prosody of synthetic speech still has high priority on the research agenda. The challenge has been summarized as follows:

> *Many researchers agree that the major remaining obstacle to fully acceptable synthetic speech is that it continues to be insufficiently natural. Progress at the segmental level, which involves the perceptually acceptable rendering of individual segments and how they conjoin, has been very successful, but prosody is the focus of concern at the moment: the rendering of suprasegmental phenomena – elements that span multiple segments – is less than satisfactory and appears to be the primary source of perceptual unease. (Tatham & Morton 2005: 2).*

The synthesis of expressive speech which conveys both emotions (cf. Chapter 13.4.2) and personality (cf. Chapter 13.5) in a believable way is another insufficiently addressed issue. Machines also need to improve their ability to adopt the appropriate tone of voice for messages with different emotional content because inconsistent verbal and vocal emotions suggest insincerity, instability, or deception on the part of the speaker (cf. Chapter 10.1.1). Synthetic speech may also sound odd if it does not follow cultural norms determining the pronunciation of lengthy numbers and other items (Nass & Brave 2005: 134f.). Machines that can produce their own messages have additional problems related to making appropriate choices with respect to content selection and organization, grammar, vocabulary, and style (cf. Section 11.1.4).

While most current synthesizers use a text-to-speech approach, more complex applications, such as speech-to-speech translation (Wahlster 2000), spoken dialogue systems (McTear 2004), and (embodied) conversational agents (cf. Section 11.3), will require the system to generate its own messages and therefore will need to use a concept-to-speech rather than a text-to-speech approach to speech synthesis.

11.1.3 Natural Language Understanding

The term *natural language understanding (NLU)* refers to the process of translating a natural language input into a non-linguistic representation of the information that the machine was able to extract from the input about its content (or meaning) and the underlying communicative goals of the speaker or writer. Natural language input is provided in spoken or written form in one or multiple languages as chunks ranging in size from single utterances to multi-sentence texts and can occur in both interactive and non-interactive settings. The concrete form of the internal representation depends on the requirements of the application. Conversational agents must be able to capture the content and function of the user's utterance at the stage at which it occurs and interpret it with respect to previous exchanges and the goal of the interaction (cf. Section 11.3).

11.1.3.1 The NLU Problem

As in other areas of speech and language processing, humans serve as the model for NLU research and development, the goal being systems with a language understanding capacity which is at least as comprehensive, flexible, and robust as the human mind. However, the interpretation of natural language input is a complex task, due to a number of inherent properties of human language:

- *Ambiguity*. One of the most serious issues for NLU is that units and constructions in natural languages are often ambiguous, i.e. several linguistic structures can be computed for them, which correspond to alternative interpretations. These ambiguities need to be resolved (*disambiguation*) (Jurafsky & Martin 2000: 4f.).
- *Vagueness*. Many expressions in natural languages are vague, i.e. *unspecified* with regard to their meaning. For example, their applicability to a particular object or situation may not be clear or their meaning may not be clearly or only generally specified. Contextual information can help to clarify vague expressions.
- *Figurative language*. Literature, rhetoric, and especially everyday language are rich in metaphor (Lakoff & Johnson 1980), metonymy, idioms, and other expressions whose meaning is different from the sum of the meanings of their parts. To be able to interpret them properly (i.e. not literally), knowledge of many "categories of shared human experience" is required (Kurzweil 1990: 307), including common sense, current events, history, literature, politics, etc.
- *Context dependency*. The meaning of extended *discourses* (texts and dialogues) can only be derived from the interaction of their components because the meaning of a text or exchange is more than the sum of the meanings of the individual sentences or utterances making it up (Schank & Abelson 1977: 22), and the discourse context[129] of an utterance affects its meaning (Grosz 1997: 199). In addition, natural language understanding has to consider other aspects of context, including speaker and hearer, the current situation, and the real world.

[129] The *discourse context* is the set of circumstances that affects the interpretation and production of a given discourse. It includes the surrounding text or talk of a language unit (word, utterance, turn, etc.) and the relevant dimensions of the communicative situation (participants, setting, and the real world).

- *Ill-formed input*. In real-world applications, particularly those involving real-time human-machine interaction, the input may contain various anomalies, including incorrect spellings, unknown and misrecognized words, speaker self-corrections, fragmentary utterances, spurious phrases, unusual and wrong word order, and missing words.

11.1.3.2 Resources for NLU

The wide range of difficulties shows that natural language understanding is a much harder problem than perhaps many people would expect. The failure of early systems soon led to the insight that understanding language requires substantial knowledge about natural languages and about the world. This knowledge can be organized in terms of five principal resources: grammar, lexicon, discourse or dialogue model, domain model, and user model.

11.1.3.2.1 Grammar

Language analysis requires knowledge about the possible organizations of words in sentences to determine the syntactic structure(s) of sentences, which is instrumental in computing their meanings. *Grammars* are finite formal specifications of the legal syntactic structures in a natural language (Charniak 1996: 4). They provide a set of basic categories and rules for combining words into well-formed phrases and sentences.

A natural language grammar is written in a formal language, which for the purposes of processing must be highly expressive, flexible, and computationally effective (Keller 1992: 378). Current grammar formalisms are typically *declarative* in nature (p. 378), i.e. they provide the means for specifying objects, their legal combinations, and the properties of the resulting objects, but do not prescribe technical procedures for the combination of objects into larger units (Gazdar & Mellish 1989). As a result, declarative formalisms become independent of particular processing strategies.

The remainder of this section identifies a number of influential grammar formalisms in the history of natural understanding and summarizes their major properties:

- *Augmented transition network (ATN) grammars* (Woods 1970; Winograd 1983) define each phrasal category (e.g. noun phrase, verb phrase, sentence, etc.) that can appear in the syntactic structure of a sentence in terms of a finite-state transition network.[130] Transitions in such a network can have phrasal category labels, which activate the ATN for that phrasal category. If a path through the subordinate network can be found, the transition arc in the parent network may be followed. Every ATN has a set of registers for storing partial syntactic structures. Operations associated with each transition arc set the values of registers and retrieve them on later transitions for comparison with items encountered further on in the input. This makes it possible to implement subject-verb agreement and other dependencies (Kaplan 2003: 84f.).
- *Context-free grammars (CFG)* (Chomsky 1956; Chomsky 1957; Backus 1959; Kaplan 2003: 78–81; Jurafsky & Martin 2008: Chapter 12) are an essential component of many computer systems for language understanding. A context-free grammar consists of (a) a finite set of *non-terminal symbols* (lexical and phrasal categories); (b) a finite set of *terminal symbols*

[130] *Finite-state transition networks* consist of states connected by links (the transitions). A transition network is traversed from an initial state through a sequence of intermediate states to a final state.

(the words of the language);[131] and (c) a finite set of *rewrite rules* (facts about sentence structure or about the lexicon) that replace a non-terminal symbol with (or rewrite it as) a string of terminal and/or non-terminal symbols. The rewriting of a non-terminal symbol is independent of the context surrounding it (hence the label 'context-free'), as opposed to *context-sensitive* grammars, where the rewriting is contingent on the context of the non-terminal symbol.

- *Unification-based grammars* (Shieber 1986; Knight 1989; Keller 1992; Kaplan 2003: 85–88; Jurafsky & Martin 2008: Chapter 15) use *feature structures* as their central type of data structure. A feature structure is a partially specified, record-like representation of linguistic objects consisting of attribute-value pairs that provide syntactic, semantic, or phonological information (Keller 1992: 371). The second central concept is the *unification operation*, which derives feature structures of larger units by putting together (unifying) compatible feature structures, which are partial specifications of the same linguistic object (p. 371). Unification-based grammars are also called *constraint-based grammars* because they impose constraints on features and values to define well-formedness criteria for linguistic constructions (p. 376), such as number agreement between subject and verb.
- *Semantic grammars* (Burton & Brown 1977; Allen 1994: 332ff.; Jurafsky & Martin 2000: 573–577) take advantage of the fact that language understanding systems are often designed for limited domains where words and phrases have well-defined meanings. The rules in a semantic grammar contain semantically motivated categories corresponding to entities and relationships of the domain and define how these entities and relationships can be expressed (Allen 1994: 333; Pereira 1997: 118; Jurafsky & Martin 2000: 575).

11.1.3.2.2 Lexicon

Lexical knowledge, i.e. knowledge about individual words in particular natural languages, is crucial for language understanding. The resource that stores lexical information in a systematic and accessible way is called the *lexicon* (Handke 1995). Computational lexicons for realistic applications are large, complex repositories of lexical information which require an effective organization, both of the global structure of the lexicon and at the level of single lexical entries (Pustejovsky 2001: 160f.). While the exact content of lexical entries is controversial, they are commonly assumed to contain at least[132] information about the following aspects (Handke 1995: 68–108 + 237–241; Pustejovsky 2001: 161):

- *Morphology*. The morphological specification contains information about the lexical category (word class), spelling, and possible word forms of a lexical item.
- *Syntax*. The syntactic specification defines the context of a lexical item in terms of other categories (*subcategorization*).
- *Semantics*. The semantic specification of a lexical item contains the basic word sense, the thematic role structure[133] of an item, selectional restrictions,[134] and finally the semantic

[131] In other accounts, lexical categories are the terminal nodes and information about words is not encoded in the grammar.

[132] Computational lexicons for spoken language processing additionally require information about the phonological aspects of words, including their sound sequence, syllable structure, stress patterns, etc.

[133] The thematic role structure displays the arguments of lexical items together with their associated thematic (semantic) roles (e.g. `Agent`, `Theme`, and `Goal`) (Handke 1995: 83).

relationships between the lexeme and other items in the lexicon (e.g. synonymy, hyponymy, and antonymy).

11.1.3.2.3 Discourse Model

The discourse model is dynamically constructed while processing an input text or in the course of a dialogue with the human user, capturing the structure of the discourse as it unfolds. Three interdependent components of discourse structure have to be captured by a discourse model (Grosz & Sidner 1986; Jurafsky & Martin 2000: 744f.):

- The *linguistic structure* organizes the discourse into a hierarchy of segments. Each segment is a coherent stretch of discourse consisting of one or more (not necessarily adjacent) utterances.
- The *intentional structure* models the purposes (or intentions) of the participants in different discourse segments and in the overall discourse.
- The *attentional state* is a dynamically updated record of the objects, properties, and relations that are salient (in focus) at any given point in the discourse.

In interactive settings, the discourse model is referred to as the *dialogue model*. More information about dialogue modeling is provided in Section 11.3.3.2.2.

11.1.3.2.4 Domain Model

Sophisticated language processing presupposes a great deal of knowledge about the context, the participants, and the world. Some aspects of this knowledge may be derived during processing, but many others have to be encoded beforehand, in particular those that represent general or expert background knowledge about the world. The component which models the information that a system has about its world (or *domain*) is called the *knowledge base* or *domain model*. Domain modeling for agents is discussed in Chapter 12.1.

11.1.3.2.5 User Model

For machines that participate in a natural language dialogue with a human user (cf. Section 11.3), it becomes crucial to maintain a *user model*, which contains explicit assumptions about all aspects of the user (cf. Chapter 6.1) that may help the system to understand the user better and to adapt its own behavior to him or her in non-trivial ways (Wahlster 1988: 102; Jameson 2003: 305). User modeling in general and for educational purposes in particular (commonly referred to as *student modeling*) are covered in Chapter 12.2.

---

[134] Selectional restrictions constrain the range of possible fillers for the argument positions of a lexical item by specifying each argument in terms of semantic features (e.g. `+animate`, `+visible`, `-male`) (Handke 1995: 83; Jurafsky & Martin 2000: 614).

11.1.3.3 The NLU Process

Machines approach the problem of natural language understanding by passing the input through a multi-stage process that draws on the linguistic and non-linguistic resources outlined in the previous section. While the nature and arrangement of stages varies across applications, it is possible to identify a number of component processes that are found in many systems. These processes can be integrated into a five-stage architecture for natural language understanding, which is shown in Figure 44. The stages of this architecture are summarized below:

- *Morphological analysis*. After the necessary preprocessing steps (speech recognition (cf. Section 11.1.1), spelling correction, text segmentation and normalization, etc.) have been performed, producing a string of words for each utterance in the input, the NLU process first has to derive linguistic information about each word-form (word class, number, gender, tense, aspect, etc.) by decomposing word-forms into their component parts and/or looking up (parts of) word-forms in the lexicon (cf. Section 11.1.3.2.2). The most widely used formalization to relate the surface level of the actual word-forms (e.g. *carries*) to the level of lexical representations (e.g. `CARRY+S`) makes use of *finite-state transition networks (FSTN)*[135] (Karttunen 2003; Trost 2003: 39–44; Jurafsky & Martin 2008: 26–41 + Chapter 3). Phonological or graphemic alternation is handled by an extension to this approach referred to as *two-level morphology* (Koskenniemi 1983; Karttunen 1983; Beesley & Karttunen 2003; Karttunen & Beesley 2005).
- *Syntactic analysis*. The task of the syntactic analysis stage is to assign a structural description to each morphologically analyzed word string of the input, using the lexical and syntactic knowledge provided by the grammar (cf. Section 11.1.3.2.1) and the lexicon in efficient algorithms[136] to analyze the strings. The process of recognizing an input string and building some kind of structure for it is commonly referred to as *parsing* (Jurafsky & Martin 2000: 57);[137] its output is a hierarchical, labeled structure showing the decomposition of the string of words into different syntactic constituents, such as `noun_phrase` or `verb_phrase`, which is called a *parse tree* (Cawsey 1998: 107 + 111). The level of detail required for the syntactic analysis depends on the task and the particular approach chosen, ranging from partial or 'shallow' to complete or 'deep' parses (Charniak 1996; Cole et al. 1997: Chapter 3; Carroll 2003; Coleman 2005: Chapters 8–9; Jurafsky & Martin 2008: Chapters 13–15):
    - *Shallow parsing*. This approach tries to extract enough information from the input to serve the purposes of the application, even if this means producing only a partial analysis.
    - *Dependency parsing*. Dependency parsing schemes derive dependency structures within sentences. The assumption is that in a syntactic relationship between two elements one is the governing, the other the dependent element.
    - *Context-free parsing*. Parsing with context-free grammars requires the parsing process to search through all possible structures generated by the rules of the grammar to find one

---

[135] In the literature, finite-state transition networks are also known by various other names, including *finite-state automaton (FSA)* and *finite-state machine*, among others. In morphological parsing, the term *finite-state transducer (FST)* is often used (Jurafsky & Martin 2008: Chapter 3).

[136] An *algorithm* consists of a finite, logical sequence of well-defined steps or instructions that, when completely executed by a human or a computer, solves a specific problem in a finite amount of time.

[137] In some approaches, the term 'parsing' takes a wider scope, also subsuming morphological analysis or semantic interpretation. Sometimes even the entire NLU process is referred to as 'parsing.'

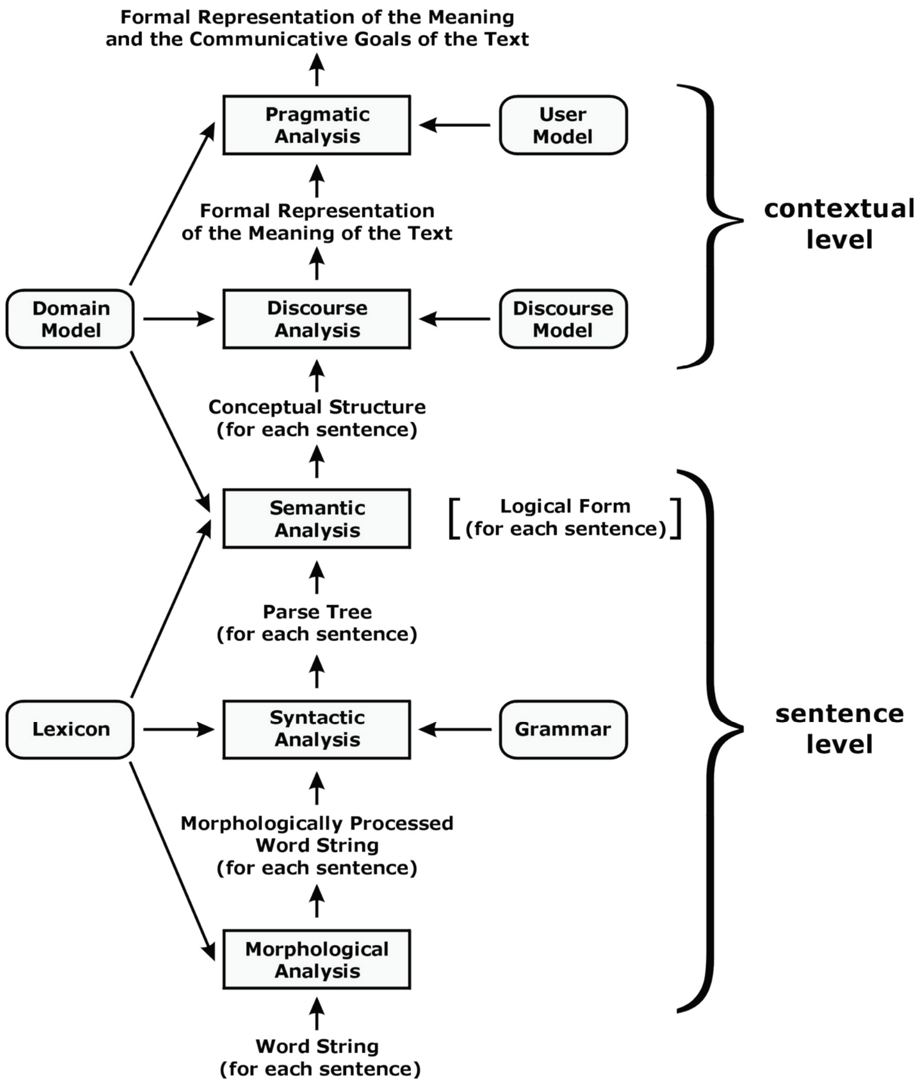

Figure 44. The architecture of a complete generic natural language understanding system. Adapted from Barros and Robin (1997).

(or several) fitting the input string (Cawsey 1998: 110), starting from the words (*bottom-up*) or with the most general rule of the grammar (*top-down*) and using *backtracking* and *look-ahead* techniques[138] to mitigate the effects of incorrect decisions.

- *Chart parsing*. Chart parsers aim to increase the efficiency of the parsing process by storing intermediate results in a well-formed substring table (the *chart*), thus making them available for later reference and reducing the need for reanalysis (Charniak 1996: 9–16; Carroll 2003: 236).
- *Unification-based parsing*. This approach represents linguistic information uniformly in terms of feature structures consisting of attribute-value pairs. Feature structures of larger units are derived by merging compatible feature structures using the unification operation (cf. Section 11.1.3.2.1).

[138] *Backtracking* is a technique that involves remembering alternative analysis paths and going back to try a different path after running into a dead-end. *Look-ahead* means that an extended context within the word string ahead of the current position is considered before making a decision.

  - *Probabilistic parsing*. These parsing algorithms use *probabilistic context-free grammars (PCFG)*, which augment every rule in the grammar with the probability of its application (estimated from a corpus of hand-parsed text called a *treebank*) and compute the product of the probabilities of the rules in a parse tree to identify the parse with the highest overall probability (Charniak 1997; Jurafsky & Martin 2000: 448–453).

- *Semantic analysis*. The main task of semantic analysis is to provide a precise account of the meaning of each sentence (Handke 1995: 30) in the input in terms of a representation in a formal language. First, the syntactic description of the sentence is mapped into a *logical form*, which contains information from the sentence structure as well as the linguistic meanings of the individual words and how they relate to one another in the sentence (Allen 1994: 14). The second stage of semantic analysis involves translating the logical form into a *conceptual structure*, which expresses the meaning of the sentence in a way that is language-independent and allows the machine to make inferences which are not possible on purely linguistic-semantic grounds (Handke 1995: 30ff.).
- *Discourse analysis.* This stage translates the set of conceptual sentence structures obtained from semantic analysis into a more comprehensive formal representation of the meaning of the input text or dialogue. This involves three subtasks:

  - Building a *model* of the unfolding discourse (cf. Section 11.3.3.2.2);
  - Resolving the *referents* of pronouns and other referring expressions;
  - Making *bridging inferences* to link explicitly mentioned discourse entities (objects, states, or events) via implicit entities (Smith 1991: 380ff.).

  Discourse analysis maintains three interrelated data structures: a stack of discourse segments that represents the linguistic structure, a stack of focus spaces for modeling the attentional state, and a script or plan[139] to capture the intentional structure reflecting the developing understanding of the discourse (Grosz & Sidner 1986; Allen 1994: Chapter 16, cf. Section 11.1.3.2.3).
- *Pragmatic analysis.* The final stage of the NLU process interprets the input with respect to its interpersonal (or participatory) context consisting of speaker/writer and hearer/reader. The output of this stage is a formal structure that captures the complete meaning of the input, including in particular the communicative goals (or intentions) of the speaker. Interpreting the intention behind an utterance involves identifying the *speech act* (Austin 1962; Searle 1979) performed with this utterance (e.g. request, warning, or promise). The interpretation of speech acts is often viewed as an application of plan recognition (cf. Section 11.3.3.2.2 and Chapter 12.2.1).

11.1.3.4 Approaches to NLU

To develop the processing steps and resources for natural language understanding outlined in the previous sections, many different techniques have been developed over the years, as documented in several comprehensive reviews and textbooks (e.g. Allen 1994; Cole et al. 1997; Jurafsky & Martin 2000; Mitkov 2003; Jurafsky & Martin 2008). The various techniques differ in the extent to which they try to solve the NLU problem (cf. Section 11.1.3.1) by mimicking human language processing or by adopting cognitively less plausible but computationally more

[139] Scripts and plans are discussed in Section 11.3.3.2.2, in Chapter 12.1.2.3, and in Chapter 12.2.1.

feasible strategies. Recent years have seen a considerable increase of interest in techniques of the latter kind, which take an engineering view of the NLU problem, focusing on the results rather than the process of natural language understanding and computing various kinds of statistics from large amounts of linguistic data to achieve those results, as discussed below (e.g. Charniak 1996; Manning & Schütze 1999; Coleman 2005).

Allen (2001) discussed three broad classes of approaches to natural language understanding, which are summarized below. He noted that these categories are not mutually exclusive; in fact, the most comprehensive models for NLU combine techniques from all three categories to take advantage of their relative strengths (Allan 2001: 593):

- *Statistical approaches* achieve language understanding by relying on information extracted from large corpora, i.e. bodies of machine-readable (annotated) linguistic data, both spoken and written. Typically, the information extracted from a corpus consists of statistics and is often processed further into rules, lexical entries, and various other forms (Charniak 2001: 801). An important feature of statistical models is that they can be acquired (trained) automatically from a corpus, whereas the rules of structural frameworks have to be crafted by hand. Statistical NLU systems can handle a wide range of inputs, in particular incomplete or incorrect ones, but their analyses tend to be shallow rather than deep and may be less than completely accurate (p. 801). However, statistical methods have already proven effective in several tasks, such as speech recognition (Jelinek 1998), syntactic parsing (Charniak 1997), word-sense disambiguation (Charniak 1996: Chapter 10), and machine translation ([INT 85]).
- *Structural approaches* to natural language understanding are based on formal models of the structure of natural languages. They are concerned with the design of expressive grammar formalisms with good computational properties, and of efficient parsing algorithms that assign structural descriptions to sentences according to a grammar (cf. Section 11.1.3.3). Many practical systems have used semantic rather than syntactic grammars (cf. Section 11.1.3.2.1), exploiting the fact that in their (often limited) domain, words and phrases have well-defined meanings. Other systems forego a complete analysis of sentence structure and instead attempt to match patterns of lexical, syntactic, and semantic information against fragments of the input. Overall, structural models can be used to produce in-depth analyses of natural language inputs, but deeper levels of analysis can only be achieved with hand-crafted rules rather than with rules automatically trained from a corpus.
- *Reasoning-based approaches* view the interpretation of natural language as a highly context-dependent process. Hence, they equip computer programs with knowledge available to humans in different situations and use this knowledge in reasoning processes to address problems of contextual interpretation, including disambiguation of words and sentences, resolution of referring expressions, and identification of speakers' or writers' intentions underlying their contributions. Plan-based techniques play a key role in reasoning-based NLU systems. While knowledge-based reasoning allows the interpretation of language in context, the (computational) complexity and limited coverage of the models involved is a major shortcoming of this approach.

11.1.3.5 State of the Art

NLU technology has reached a level of development that allows its use in a small but growing number of limited practical applications. But complete understanding of unrestricted natural language input remains elusive even after more than half a century of intense research and development. Success stories involve either very small domains or relatively easy tasks. Many programs aim to support human expertise rather than attempt to replace it with a fully automatic solution. Applications where techniques for natural language understanding are already useful include authoring aids (spelling correction, thesauri, hyphenation, grammar and style checking), information extraction (Cowie & Wilks 2000; Grishman 2003; Cunningham 2006; Sarawagi 2008), information retrieval (Baeza-Yates & Ribeiro-Neto 1999; Tzoukermann et al. 2003; Grossman & Frieder 2004), question answering (Harabagiu & Moldovan 2003; Maybury 2004), text categorization and summarization (Mani 2001; Hovy 2003; Das & Martins 2007), machine translation (Hutchins & Somers 1992; Arnold et al. 1994; Trujillo 1999), intelligent computer-assisted language learning (ICALL) (Nerbonne 2003), natural language database interfaces (Androutsopoulos & Aretoulaki 2003: 630–635), spoken dialogue systems (McTear 2004), chatbots (cf. Section 11.3.2), dialogue agents (cf. Section 11.3.3), and embodied conversational agents (cf. Section 11.3.4).

While many early NLU programs were ‘toy systems’ with very limited coverage of natural language phenomena (Gazdar 1993: 164), in the early 1990s the field began to move from restricted experiments to large-scale, practical, and evaluable applications, shifting the emphasis from academics to language engineering. The most progress has been made in the areas of morphological and syntactic analysis, where ready-to-use techniques and software are now available. Furthermore, there are a number of *engines* for natural language interpretation that generate semantic analyses of sentences (e.g. Alshawi 1992). However, substantial problems persist at the contextual level of the NLU process (cf. Figure 44). There is yet no program that can interpret a discourse and the goals of its participants in a sophisticated way. Future research not only needs to close many remaining gaps in our scientific understanding of natural language but also has to address practical issues:

- *Robustness*. Develop techniques that maintain good performance when faced with distorted or unknown input.
- *Broad coverage*. Find ways to overcome the knowledge acquisition bottleneck (created by the need to code knowledge for natural language understanding by hand) and build large-scale linguistic and non-linguistic resources for NLU.
- *Reference architecture*. Build a generic framework for NLU systems that can serve as a reference for research and development.
- *Evaluation*. Provide general tools and frameworks to assess the performance of NLU systems (Sparck Jones & Galliers 1996).

11.1.4 Natural Language Generation

Natural language generation (NLG) is the computer-based process of constructing outputs in natural language from inputs provided in some kind of non-linguistic format in order to achieve specified communicative goals (McDonald 1992; Reiter & Dale 1997; Jurafsky & Martin 2000: Chapter 20; Reiter & Dale 2000; Hovy 2001; Bateman 2002; Bateman & Zock 2003). In other words, the problem of natural language generation is to enable computers to express their own messages using appropriate linguistic means. NLG produces messages ranging in length from single phrases or sentences through multi-sentence paragraphs to whole documents consisting of multiple paragraphs. The messages originate in both interactive and non-interactive settings, are generated in spoken or written form in one or several languages, and may be aimed at an individual user or a broader target audience.

11.1.4.1 The NLG Problem

Systems for natural language generation should be able to produce linguistic outputs that are meaningful, eloquent, and contextually appropriate. The traditional approach to the problem involves two major tasks: *content selection* (determining *what* to say) and *content expression* (finding a way *how* to say it) (Hovy 2001: 589). Both tasks are by no means trivial, given that natural languages have the expressive power to communicate information in many different ways (Cole et al. 1995). Hence, the machine has to make choices at many levels (Jurafsky & Martin 2000: 766), from planning the overall structure of the output down to finding an appropriate word or phrase to express a specific concept. The major problems involved in the generation of natural language output are summarized below (Jurafsky & Martin 2000: 765f.; Hovy 2001):

- *The nature of the input*. NLG systems for different applications need to deal with a wide variety of input specifications (Jurafsky & Martin 2000: 765), including such diverse sources as numerical data, results of database queries, formal models, log files, images, charts, graphs, meaning representations produced by NLU (cf. Section 11.1.3), and knowledge bases. The format of these representations is often application-specific, but they are also more structured, less ambiguous, and more explicit compared to the linguistic input of natural language understanding systems (p. 766).
- *Content selection*. The generation system must decide which of the information available to present in the output. Content selection involves a number of difficult tasks, such as picking only the relevant content items, determining the amount of detail required, uncovering and filling gaps in the input, and identifying material that can be inferred from the context and thus need not be mentioned explicitly. Presenting the right amount of relevant information to the user is essential. Therefore, NLG systems require at least some default assumptions about the previous knowledge and expectations of the addressees of their outputs, or, in applications involving regular or longer interactions with particular people, a dynamically constructed model of the individual user (cf. Section 11.3.3.2.3).
- *Lexical selection*. NLG systems require the ability to make appropriate choices of words and phrases in a given context of generation (Jurafsky & Martin 2000: 790). Lexical selection is influenced by many interacting linguistic and non-linguistic factors (Stede 1996), including the discourse history and context, the language level, the intended stylistic effect, collocational and selectional restrictions,[140] dialectal variation, and the beliefs, expertise, and views of the user (cf. Chapter 12.2.1).
- *Sentence structure*. At the level of the individual sentences making up the output, the generation system has to make many decisions concerning the number and length of sentences and their sequence, the information to express in each sentence and the organization of the material inside a sentence, the generation of referring expressions (pronouns, proper names, etc.) and discourse markers (e.g. *all in all*, *consequently*, *however*, etc.), and the specification of syntactic parameters (tense, mood, sentence type, etc.) for each sentence (Hovy 2001: 590).
- *Discourse structure*. The structure of the natural language output has to be carefully planned to create a coherent and fluent text or dialogue move that communicates the information in the intended way. NLG systems must be able to identify the discourse relations between the

[140] These are restrictions on what words are conventionally or can be meaningfully used together.

different pieces of information, and they have to observe the conventions for different types of natural language output (Schmitz 1992: 186f.).
- *Stylistic consistency*. Generators with the ability to express the same thing in a variety of ways have to be guided through the multiple options that may be available at the different stages of the generation process to ensure that consistent choices of words and constructions are made in order for the output to have a uniform and appropriate style (Hovy 1997: 144; Hovy 2001: 591; Bateman & Zock 2003: 300).

11.1.4.2 The NLG Process

Natural language generation cannot be accomplished without substantial knowledge of different kinds (Reiter et al. 2003: 493), including knowledge of the application domain (content material, relevance of individual pieces of information, and connections between information units), knowledge of the target language for generation (discourse, grammar, lexicon, morphology, and orthography), rhetorical knowledge (communication strategies, presentation techniques, text types, style, etc.), engineering knowledge about how to break down, represent, and organize the NLG process, and finally knowledge of the target audience of users (cf. Chapter 12.2.1) (Bateman & Zock 2003: 285). Natural language generation has been described as a planning task (Hovy & Wanner 1996), a process of progressive refinement (McDonald 1992: 984), which transforms a communicative goal and non-linguistic information into an output in a particular natural language.

The modern view of the natural language generation process is a three-stage pipeline (Reiter & Dale 1997) consisting of discourse planning, sentence planning, and surface realization, all three being guided by a stylistic control module (cf. Figure 45). The components of this architecture are discussed below (Reiter & Dale 1997; Hovy 2001):

- *Discourse planning*. This stage accepts a communicative goal and information from the host program[141] and performs the following language-independent operations (Reiter & Dale 1997: 9f.; Hovy 2001: 590):
  - *Content selection*. Choose the information to be expressed in the output, by filtering, summarizing, and otherwise processing the input data.
  - *Discourse structuring*. Create a coherent structure for the selected content by determining the conceptual groupings and discourse relationships among the items (p. 17).

  The output of this process is a fully specified *discourse plan* that reflects all the choices made about the content and structure of the planned communication as a whole by specifying the content items to be expressed and how they are grouped together and related to each other (Reiter & Dale 1997: 17; Jurafsky & Martin 2000: 767).
- *Sentence planning*. The sentence planner converts the discourse plan into a sequence of *sentence plans*, each of which corresponds to a sentence in the output. It combines (aggregates) appropriate content items into phrases, clauses, and sentences, choosing from a set of syntactic mechanisms including conjunction, ellipsis, set formation, and embedding (Reiter & Dale 1997: 20f.). Furthermore, the sentence planner selects specific words and phrases for expressing concepts, events, and the relationships between them and for referring

[141] The *host program* is an external process that initiates the NLG process, e.g. a database management system, a machine translation engine, or the dialogue manager of a conversational agent.

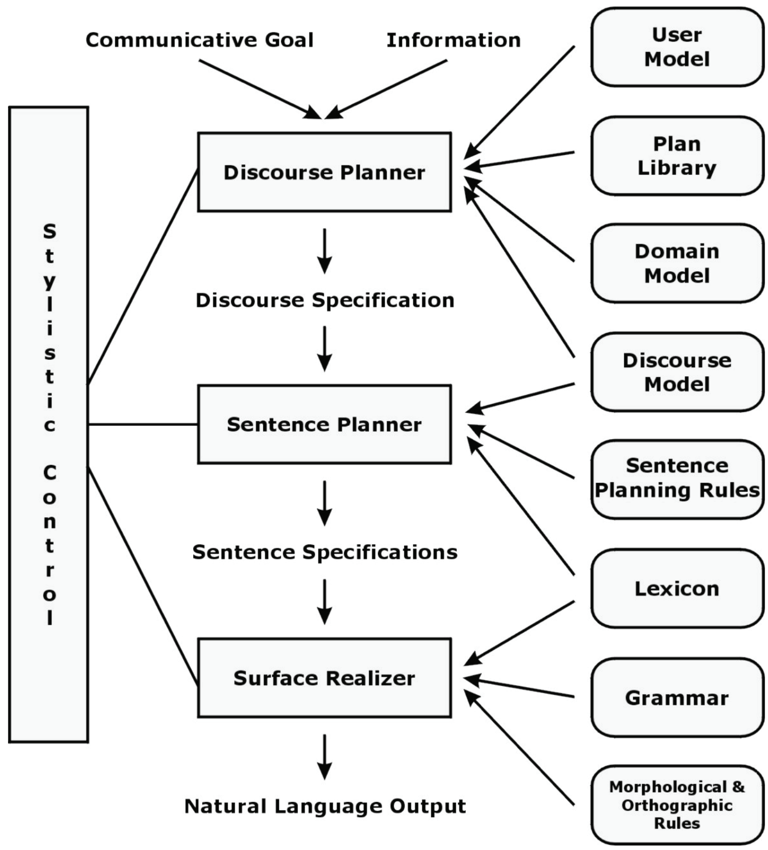

Figure 45. The architecture of a complete generic natural language generation system. Adapted from Hovy (2001: 590, Figure 1).

to the entities being discussed (pp. 11f.). In addition, this stage makes various decisions about the structure of the output sentence, which include assigning syntactic functions (subject, object, etc.) to constituents, determining the clause structure, and specifying syntactic parameters like tense, mood, active or passive voice, etc. (Hovy 2001: 590). Finally, the sentence planner devises ways to convey discourse relationships to other sentences (e.g. by using discourse markers).

- *Surface realization*. This stage transforms each sentence plan into a sequence of words which is syntactically, morphologically, and orthographically well-formed, as constrained by the lexicon, the grammar, and the set of morphological and orthographic rules of the target language (Reiter & Dale 1997; Jurafsky & Martin 2000: 767). In particular, it determines word order, function words, inflections, punctuation, upper and lower case, representation of numbers, and so on (Hovy 2001: 590).
- *Stylistic control*. Stylistic control guides the generation process from beginning to end to ensure that the choices made at the individual processing levels yield a natural language output with a consistent style. The preferred option at each choice point in the generation process is determined by a pre-defined set of pragmatic parameters for the application specifying the degree of formality, time frame, user language level, genre, etc. of the communication (Hovy 2001: 591).

11.1.4.3 Generation Techniques

To perform the tasks of the generation process, many different techniques have been developed that range from the inflexible and simplistic to the extremely flexible and sophisticated, most of them for surface realization. However, almost all NLG systems have used at least one of the following basic generation techniques (Reiter & Dale 1997; Jurafsky & Martin 2000: Chapter 20; Hovy 2001):

- *Canned text*. In very simple applications, the generation system selects pre-defined strings of words and delivers them unchanged. Canned output may consist of single or multiple sentences. While simple to implement, the chief disadvantage of this method is lack of flexibility: because the messages are fixed, new situations or languages require the extension or modification of the system's string database by a programmer (Jurafsky & Martin 2000: 763).
- *Templates*. More flexible than canned text are systems that use pre-defined structures which combine static and variable components (so-called *templates*). The variable parts of a template are blank spaces (*slots*) whose fillers are specified by the content to be expressed (Hovy 2001: 591). Slots can be associated with constraints on the information that may fill them. Furthermore, the slot-fillers may be modified by additional linguistic processing to achieve number agreement or inflection, for example (Reiter & Dale 1997). While lack of flexibility is a shortcoming that applies to templates as well (though less so than to canned text), templates have been successfully used for both planning and realization tasks. Chatbots (cf. Section 11.3.2), for example, generate their answers by filling templates with pieces of information which they extract from user inputs. Furthermore, it has been argued that recent template-based NLG systems actually perform better than the general bad reputation of these systems suggests (van Deemter et al. 2005).
- *Cascaded patterns*. This approach starts from an abstract pattern that matches the input and recursively expands this pattern by replacing its components with more specific patterns, which are selected according to the content to be conveyed. The cascading process stops when the desired target elements have been reached. The resulting hierarchy is conveniently represented as a tree structure (Hovy 2001: 591). Cascaded patterns have applications in surface realization, where rules of the grammar serve as patterns for creating syntactic tree structures whose leaves are the words of the sentence (p. 591). In discourse planning, cascaded patterns are used to generate the overall structure of the natural language output (Reiter & Dale 1997).
- *Text schemata*. Text schemata (McKeown 1985) are pre-defined patterns detailing how discourse plans can be built from smaller schemata or atomic units of content that are linked by discourse relations (e.g. contrast, elaboration, purpose, etc.) (Reiter & Dale 1997). Planning the structure of the discourse is then a matter of selecting and combining the required schemata, and of filling in the information to be presented. Text schemata are the simplest and most popular discourse structuring method. They have proven useful for producing texts in limited domains, which follow a small number of consistent structural patterns. While the advantage of using text schemata is in their simplicity, the fixed structure of the schemata makes them unsuitable for generating outputs with more varied expressions and less stereotypical structures (Jurafsky & Martin 2000: 782).

  More flexibility than by pre-defined text schemata is offered by approaches that make use of planning methods developed in the field of artificial intelligence to dynamically construct discourse plans by recursively decomposing some top-level communicative goal into more specific subgoals, using operationalized discourse relations as plan operators (Hovy 1993; Jurafsky & Martin 2000: 785–788; Bateman & Zock 2003: 292–296, cf. Section 11.3.3.2.2). The resulting plan is both a representation of discourse structure and a plan of action for achieving the communicative goal (Hovy 1993: 355). The set of discourse relations provided by *Rhetorical Structure Theory (RST)* (Mann & Thompson 1988) has been widely adopted in the NLG community (Reiter & Dale 1997). RST defines a set of 23 *rhetorical relations* that link a central segment of text (the *nucleus*) to a more peripheral segment (the *satellite*), e.g. ELABORATION and PURPOSE, or connect several nuclei, e.g. CONTRAST and SEQUENCE. RST relations combine into a hierarchy to describe the rhetorical structure of a given text (Jurafsky & Martin 2000: 782–785). Reiter and Dale (1997) pointed out that at present, scientific understanding of the different discourse relations, their precise effects, and the conditions of their application is fairly limited. Furthermore, planning-based approaches are computationally expensive and require extensive knowledge. As a result, they are not as popular in current practical NLG systems as text schemata (Reiter & Dale 1997).

- *Features*. Feature-based systems are the currently most advanced approach to surface realization. Based on the simple idea that "any distinction in language can be added to the system as a feature" (Hovy 1997: 134), they uniformly represent the generation grammar, the lexicon, and the input specification as feature structures. Each sentence to be produced is characterized in terms of a distinct collection of features (e.g. POSITIVE, STATEMENT, PAST, etc.), which are eventually transformed into a sequence of words. Feature specifications, or *functional descriptions (FDs)*, of sentences are built incrementally through *unification* (cf. Section 11.1.3.2.1). The resulting, fully unified FDs are then used to produce a grammatical string of words in each case (Jurafsky & Martin 2000: 774–778). While feature-based systems are built on a simple yet powerful concept, it is difficult to maintain relationships between features and to control their selection (a larger number of available features means higher input complexity) (Hovy 1997: 134). To date, feature-based techniques have been used only rarely for discourse or sentence planning.

11.1.4.4 State of the Art

Natural language generation technology enables computers to convey information to users in a way that is familiar and easily comprehensible to them (Reiter & Dale 1997). Current practical NLG systems can generate outputs of usable quality from a variety of input specifications for tasks such as summarization of database or spreadsheet content, production of customized information, explanation of expert system behavior and rules, question answering about objects in a knowledge base, authoring of routine letters and reports, crafting textual descriptions of graphical presentations of data, automatic generation of multimedia documents and presentations, and building interactive game characters and self-explaining virtual worlds. For an older survey of applied NLG systems, see Paiva (1998). Theune (2003) reviewed natural language generators with respect to their suitability for dialogue-based systems.

However, current systems operate successfully only in small, well-defined domains and possess limited expressive capabilities, which are also often tuned to the needs of specific applications. The most advanced solutions exist for surface realization, on which much work in the field seems to have focused. But generation grammars and lexicons are still too small to be useful for large-scale, real-world applications. Sentence planners and discourse planners are even more restricted. In general, knowledge acquisition for natural language generation systems, whether manually by knowledge engineers or automatically using machine learning techniques (cf. Chapter 12.1.1), remains a difficult challenge, with the need to ensure consistent quality of the knowledge acquired being the major problem to be solved (Reiter et al. 2003).

In summary, natural language generation technology must be developed further to be useful in realistic domains (Hovy 1997: 144). The following issues need to be addressed by ongoing research:

- *Robustness*. Develop techniques that enable natural language generators to operate well even when information is inaccurate or missing.
- *Improved output quality*. Enable NLG systems to produce more fluent, natural, and varied output that comes closer in quality to human authoring.
- *Broad coverage*. Build large-scale, well-structured linguistic and non-linguistic resources for NLG, including lexicons, grammars, general-purpose knowledge bases, and plan libraries (Hovy 1997: 144f.; Reiter et al. 2003).
- *Utterance vs. text generation*. Develop generators that do not produce longer texts like most traditional NLG systems (Theune 2003: 3) but rather concise contributions to a spoken dialogue that are contextually, culturally and individually appropriate (cf. Section 11.3.3.3).
- *Reference architecture*. Build an "end-to-end" model of the NLG process, which comprises a set of principal processing modules and data interfaces between modules and can serve as a reference for research and development. For a recent proposal, see the report by Mellish et al. (2006).
- *Evaluation*. Provide general tools and frameworks to assess the performance of natural language generation systems.

11.2 Multimodal Interfaces

The problem, in my opinion, is that our current computers are both deaf and blind: they experience the world only by way of a keyboard and mouse. Even multimedia machines, those that handle audiovisual signals as well as text, simply transport strings of data. They do not understand the meaning behind the characters, sounds and pictures they convey. I believe computers must be able to see and hear what we do before they can prove truly helpful. What is more, they must be able to recognize who we are and, as much as another person or even a dog would, make sense of what we are thinking. (Pentland 1996: 54).

While natural language is one of the most important vehicles for humans to express their thoughts, emotions, and intentions and to learn about those of others, communication is not restricted to the use of language. In fact, communication between people combines language with facial, gestural, and a variety of other cues to convey messages (Maybury 2002). Humans naturally make use of multiple channels or *modalities*[142] (visual, auditory, tactile, gustatory,

[142] Three concepts are fundamental to an understanding of multimodal interfaces: medium, modality, and mode. The term *medium* denotes a container for information to be transmitted, presented, or stored ([INT 86]). Examples of media include text, video, graphics, animation, and audio. A *modality* is a communication path between a human being and a computer, which is characterized by a human sense (vision, audition, tactition, gustation, olfaction, etc.) for receiving output of the computer and a corresponding sensor or device (e.g. camera, microphone, pen, keyboard, mouse, heat or pressure sensors) of the computer for receiving input from the user ([INT 87]). The *mode* of communication is the means by which messages are transmitted between communication partners through a modality. Examples include speech, writing, facial expressions, gestures, and touch.

olfactory, etc.) through which information may pass between the participants in an interaction. The information communicated through the channels is partly complementary (completing the information provided by the other channels), partly redundant (providing the same information in a different format), or it may contribute new items to the interaction that are not conveyed by the other channels. The resulting increased communication bandwidth allows people to transmit and process more information, to use information provided by one modality to disambiguate inputs from other modalities (e.g. watch the interlocutor's lip movements to identify words spoken against background noise or follow a deictic gesture (cf. Chapter 10.1.2) to pick out the referent of a pronoun), and to select the contextually most appropriate way to express certain information (e.g. by speaking, gesturing, or drawing).

Multimodal user interfaces (Maybury 1993; Waibel et al. 1996; Cassell et al. 2000b; Oviatt & Cohen 2000; Turk 2001; Maybury 2002; Oviatt 2003; Jaimes & Sebe 2005; Pantic et al. 2007) aim to emulate the ways in which humans make use of multiple communication channels (modalities) in their interactions with other people, with the goal of increasing the efficiency, naturalness, and robustness of exchanges between humans and machines. According to Oviatt (2003: 287), multimodal interfaces accept combined user actions from two or more input modes, using at least one recognition-based technology (e.g. speech, pen, or vision), process them in a coordinated way, and generate multimedia outputs to respond to the user. Depending on the degree of user awareness and involvement, input modes can be classified as active or passive (Oviatt 2003: 290).[143] *Active input modes* capture explicit commands to a multimodal system that are formulated by means of intentional user actions, such as gesturing, pointing, drawing, writing, typing, speaking, and combinations of these. In contrast, *passive input modes* involve the multimodal system unobtrusively monitoring natural user behaviors, such as facial expressions, gaze, spontaneous gestures, posture, and presence, which are not intended as commands to the machine. Passive input modes have the advantage that they can collect information without disrupting the user; however, active input modes provide more reliable information about what the user wants to achieve (p. 290). *Hybrid* or *blended multimodal interfaces* combine both active and passive input modes to exploit their respective strengths while compensating for the weaknesses of each. On the one hand, the active input modes can help to clarify the user's intention if the results of passive monitoring are ambiguous. On the other hand, advance information from a passive input mode (e.g. gaze) can help the system to predict and interpret inputs from a following active mode (e.g. speech) (Oviatt & Cohen 2000: 51). One example of such a blended multimodal system are embodied conversational agents (cf. Section 11.3.4), which accept combined inputs in the auditory and visual modalities, such as speech (active mode) accompanied by facial expressions (passive mode).

The outputs generated by a multimodal user interface may consist of combinations of images, text, speech, sound, video, and animations. In embodied conversational agent systems, the responses are delivered by an animated character (cf. Chapter 9) that performs both verbal and non-verbal behaviors, as described in Chapter 10, using its virtual body like a human being does in face-to-face conversations with other individuals.

[143] The distinction between active and passive input modes can also be made in terms of control versus awareness. *Control* means explicit communication of users with a multimodal system through active input modes, whereas *awareness* involves the use of passive input modes by the system to collect information about the user without explicit communication (Turk 2001: 45). Awareness thus adds *context* to the multimodal interface or the application (p. 45), i.e. information about the who, where, what, how, when, and why of the interaction with humans (Pantic et al. 2007: 52f.).

Figure 46 shows an abstract architecture of a multimodal user interface. Maybury (2002) described the input and output processing in this architecture as summarized in the following. User input can be obtained by means of various input devices, such as mouse, pen, keyboard, microphone, and equipment for tracking body, face and eye movements.[144] Processors for speech, facial expressions, gestures, etc. then analyze the different input modes separately. The next step is to integrate (fuse) the information extracted from the different input modes into an overall meaning representation (see below). Next, an interaction manager performs tasks such as recognizing the user's intention, planning the next move of the system, and maintaining models of the user, dialogue, context, and task. Furthermore, it interacts with one or more back-end application systems to obtain information or carry out user commands. To produce the multimedia output, the system goes through the steps of generating a coherent and consistent response that includes the items selected for communication; synthesizing the necessary media elements (speech, graphics, character performances, etc.); and finally rendering the output for various presentation technologies, such as display screens, audio devices, and animated agents. All stages of input and output processing in the architecture are informed by (and build) models for the recognition of speech, facial displays, pen-based input, etc. and for the production of audio, text, character behavior, and so on, as well as elaborate models of the user, dialogue, context, domain, task, and application systems.

Multimodal systems vary in the degree to which they intend to enable the machine to interact with people in human-like ways. One kind of multimodal interface that explicitly aims to "make human-computer interaction more like how people interact with each other and with the world" (Turk 2001: 39) are the so-called *perceptual user interfaces (PUIs)* (Turk & Robertson 2000; Turk 2001). PUIs combine perceptual, multimodal, and multimedia elements to model the modalities, mechanisms, and competencies involved in natural human-human interaction, with the goal of "creating more natural and more intuitive interfaces" (Turk & Robertson 2000: 34), freeing users from the constraints of current graphical user interfaces (GUIs) by offering them multiple ways to interact with machines. Perceptual interfaces leverage users' natural skills in perceiving their environment, as well as their early-acquired social skills that facilitate their interactions with other individuals (Turk 2001: 43).

Perceptual user interfaces require at least the ability to hear, see, and model the individual user and to respond to him or her with multimedia system output (including performances by animated characters[145]) that takes advantage of human perceptual capabilities (Turk 2001: 43). PUIs may also include further human modalities, such as sense of touch, possibly enhanced in comparison to people (e.g. infrared vision). Other sensory equipment of PUIs may collect data that is not directly accessible to human senses (e.g. data about blood pressure, muscle action potentials, perspiration, etc.), enabling perceptual user interfaces to more accurately recognize physiological indicators of different emotions, for example (cf. Chapter 13.1.1). Embodied conversational agents, to be discussed in Section 11.3.4, are one kind of perceptual user

[144] See Hinckley et al. (2004) for a review of various input and output devices as well as interaction techniques involving those devices.

[145] Given the ongoing debate about the effectiveness of anthropomorphic interfaces (cf. Chapter 9.1), one proponent of PUIs has been careful to point out that "[p]erceptual interfaces do not necessarily imply anthropomorphic interfaces" (Turk 2001: 49). However, he continued: "It is likely that, as computers are seen less as tools for specific tasks and more as part of our communication and information infrastructure, combining perceptual interfaces with anthropomorphic characteristics will become commonplace" (p. 49).

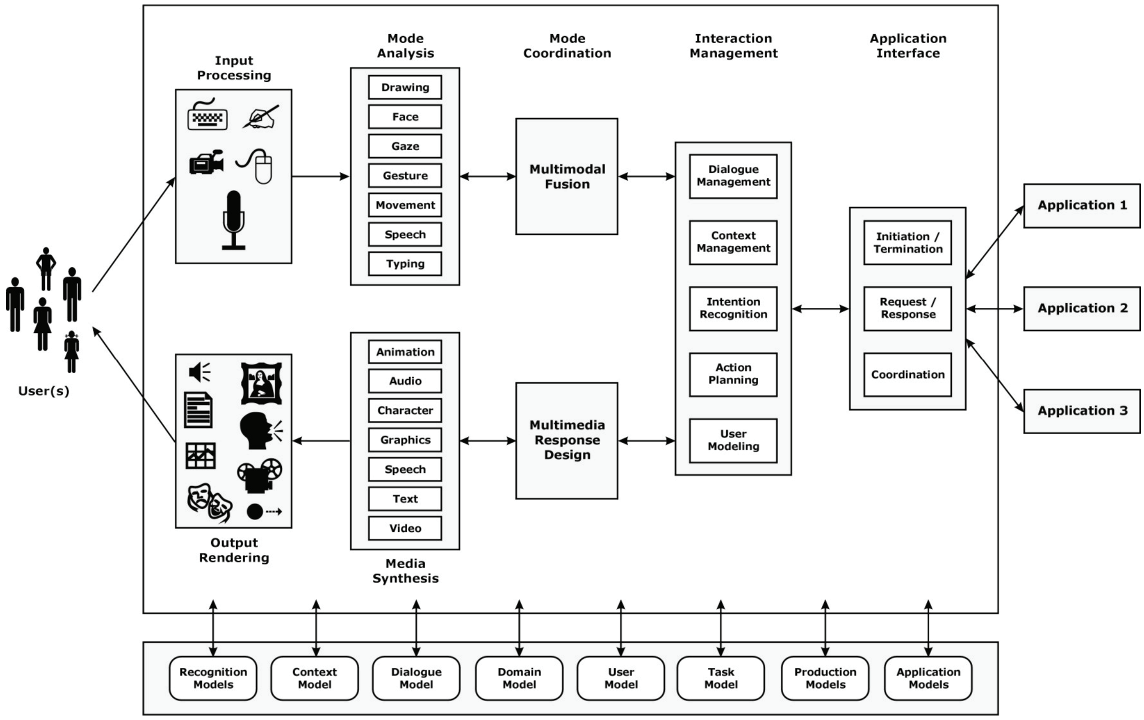

Figure 46. An abstract multimodal interface architecture. Adapted from Maybury (2002, Figure 1) and Oviatt (2003: 299, Figure 14.4).

interface, which are explicitly equipped to participate in face-to-face conversations with human beings.

The chief motivation for research and development of multimodal user interfaces lies in their potential to make interactions between humans and machines more natural, efficient, and robust. The specific advantages of multimodal interfaces over others that rely only on a single modality are summarized below (Oviatt 1999; Oviatt & Cohen 2000; Oviatt 2003: 290–293):

- Multimodal interfaces allow the flexible selection, combination and switching of input and output modes, according to each user's abilities, constraints, convenience, and preferences.
- Multimodal interaction can adapt to different user characteristics, including age, skill level, cognitive style, sensory and motor impairments, native language, temporary illnesses, and permanent handicaps, thus increasing the accessibility of computer systems for a wider range of non-expert users.
- Switching modes reduces the risk of overuse of and resulting physical damage to individual modes (e.g. repetitive strain injury from typing).
- Multimodal interaction can increase efficiency in a number of applications, especially when manipulation of graphical information is involved.
- Users choose multimodal over unimodal interaction for a wide variety of tasks, particularly in spatial domains (e.g. interaction with maps).
- In mobile applications, users can flexibly switch to a different input or output mode when environmental conditions change.
- Multimodal systems can use partial information provided by one input mode to solve problems resulting from ambiguous, imprecise, or incorrect input or processing errors in another mode. This ability to recover from unimodal recognition errors reduces the overall number of errors experienced by the user (Oviatt 2000).

- Multimodal interfaces allow for the selection, combination, and switching of input and output modes to create a synergistic blend (Oviatt 2000: 45) which takes advantage of the strengths of each mode and compensates for its weaknesses. Hence, multimodal interfaces can perform more robustly and under less favorable conditions than unimodal systems.

While this is an impressive list of arguments in favor of designing user interfaces for multimodal interaction, developers considering the integration of multiple modalities into their applications should be aware of a number of common misconceptions, or myths, about multimodal interaction. Table 15 lists ten myths of multimodal interaction (compiled by Oviatt 1999), contrary empirical evidence (the reality), and implications for the design of multimodal systems. In general, while research has identified a number of recurring multimodal interaction patterns (e.g. gesture precedes speech), when and how a particular user chooses to interact unimodally and multimodally depends on the task to be accomplished and his or her preferred interaction style (Buisine & Martin 2005).

One major challenge in multimodal interface development concerns the integration and coordination of multiple input and output modalities during the interaction. On the input side, incoming information from several input channels has to be integrated and synchronized to compute a representation that shows how the meanings contributed by each of the input modes combine into the user's intended message. The integration (or *fusion*) of modes can happen at an early or at a later point during input processing (Stork 1997b: 252f.). *Early integration* (or *early fusion*) means that the input data from the different modes are fused following minimal processing (or none at all). By contrast, in *late integration*, each input mode is processed separately in some depth first, and the resulting partial meaning representations are combined into a final interpretation (p. 253). Hence, late integration fuses information extracted from parallel input modes at a higher, semantic level, whereas early integration fuses low-level feature information (p. 252; Oviatt 2003: 298). Early feature fusion is typically used in systems that deal with closely synchronized input modes, such as speech and lip movements, whereas late semantic-level fusion processes the complementary information provided by less temporally coupled input modes like combined speech and gesture (Oviatt 2003: 298). Semantic fusion relies on a framework for creating (partial) meaning representations that is shared among the modalities and on an operation for fusing semantic information from different input modes. One approach, *frame-based integration*, recursively combines attribute-value structures ('frames') into a final meaning representation. Another applies the *unification operation* (cf. Section 11.1.3.2.1 and Section 11.1.4.3) to typed feature structures specifying the contribution of each input mode to the meaning of the multimodal input (Oviatt et al. 2000: 278f.).

Synchronization is required to accommodate patterns of temporal (co-) occurrence of input events from different modes. As Table 15 indicates, the common assumption that input signals will happen simultaneously is not supported by the reality of multimodal interaction. In fact, multimodal input often shows no temporal overlap between the modes (Oviatt 2003: 295). For example, pen-based and spoken input were found to occur sequentially about half of the time in a multimodal interface, with speech input typically lagging 1–2 seconds behind pen-based input (Oviatt et al. 1997). Even speech and lip movements lack perfect temporal synchrony, the latter preceding the former by the fraction of a second (Abry et al. 1996; Benoît 2000).

Table 15. Multimodal interaction: myths, empirical reality, and implications for the design of multimodal systems. Adapted from Oviatt (1999).

Myths	Reality	Implications
If you build a multimodal system, users will interact multimodally.	Users communicate naturally by combining unimodal and multimodal expressions.	Multimodal systems have to distinguish between users' unimodal and multimodal communication.
Speech and pointing is the dominant multimodal integration pattern.	Only a minority of multimodal interactions involve a speak-and-point integration pattern.	Other input modes and their integration patterns have to be considered.
Multimodal input involves simultaneous signals.	While two input modes may be highly interdependent and synchronized, they often do not occur together in time.	Developers should not rely on a convenient overlapping of input signals.
Speech is the primary input mode in any multimodal system that includes it.	Important information is also communicated by modes other than speech, and speech typically does not occur before other input modes.	The contributions of other input modes should not be neglected.
Multimodal language does not differ linguistically from unimodal language.	Multimodal language is different from unimodal language. In many respects, it is much simpler.	Processing multimodal language may be less difficult than processing language alone.
Multimodal integration involves redundancy of content between modes.	Multimodal inputs contain much more complementary than redundant information.	Processing of multimodal input should not assume that different input modes will provide the same information.
Individual error-prone recognition technologies combine multimodally to produce even greater unreliability.	Recognition can be more robust because (a) users select input modes to avoid or resolve errors; and (b) the synergistic blend of input modes facilitates mutual disambiguation.	Multimodal interfaces can perform better than other kinds of interfaces because of their superior error-handling capabilities resulting from the selectability and synergy of input modes.
All users' multimodal commands are integrated in a uniform way.	Users may differ considerably with respect to their preferred multimodal integration patterns.	Multimodal systems should be able to identify the prominent integration pattern of each user and adapt to it.
Different input modes are capable of transmitting comparable content.	Input modes differ with respect to the type of information that can be conveyed through them, the way they function in communication, their integration with other modes, and their suitability for different types of interfaces.	Developers have to consider the capabilities and limitations of each input mode in a multimodal system.
Enhanced efficiency is the main advantage of multimodal systems.	Multimodal systems have many other performance advantages, including reduced error rates during multimodal interaction, user preference, flexibility of mode selection and alternation, improved error avoidance and recovery, and better accessibility.	Efficiency is only one advantage of multimodal interfaces among many others and may not be the most important one for the application being designed.

For the generation of multimedia output, it is necessary to decide what media to use, what information to express in each medium, and how to combine the media into a coherent multimedia response that is contextually appropriate and tailored to the user's abilities, preferences, and task (Maybury 1995: 109–117). If the multimodal interface involves an embodied character, all the aspects related to the embodiment and behavior of animated agents that were discussed in the previous two chapters have to be considered.

Multimodal interfaces involve various component technologies which are in constant need of further development and innovation. In addition to the human language technologies reviewed in the previous section, multimodal interfaces require technologies for processing non-linguistic modes of communication. A number of technologies based on computer vision (Forsyth & Ponce 2003) which are relevant for perceptual user interfaces and embodied conversational agents (cf. Section 11.3.4) are described below:

- *Face detection* is the problem of identifying all regions in a digital image that contain a human face, regardless of clutter, occlusions, variations in head pose and lighting conditions, facial movements, and different sizes, colors, and textures of faces (Pantic et al. 2007: 54). The majority of the face detection techniques developed so far is based on statistical learning and makes use of appearance features (p. 54), like for example the popular real-time detection algorithm developed by Viola and Jones (2001). The most severe restriction of current face detection techniques is that they work reliably only for (near-) upright faces in (near-) frontal view (Pantic et al. 2007: 54).
- *Tracking* involves finding and following the changing location of one or several moving objects, in particular the face, head, hands, or body, within a video sequence captured by a camera. Image features such as points, lines, or blobs are used in an analysis of the video frames to determine the location of a moving target object within each frame of the sequence. Locating the image of a target object in successive frames is quite difficult, in particular for objects that move fast with respect to the frame rate of the video ([INT 88]). Problems arising from occlusion, clutter, and changes in lighting conditions also need to be addressed (Pantic et al. 2007: 54). Different tracking methods include *model-based tracking*, which represents the geometric structure of the human body as a stick figure (i.e. line segments connected by joints), 2D contour (ribbon or blob), or volumetric model (consisting of e.g. cylinders, cones, spheres, etc.); *region-based tracking*, which identifies connected regions in an image that correspond to moving objects and tracks these regions over time by computing their cross-correlation; *active contour-based tracking*, which extracts a representation of the bounding contour (shape) of an object and dynamically updates this representation over time; and *feature-based tracking*, which, rather than tracking objects as a whole, extracts and matches features of an object, such as distinctive points, lines, or blobs (see Wang et al. (2003) for a detailed discussion of these methods). Surveys of vision-based techniques for human motion analysis have been contributed by Gavrila (1999) and Wang et al. (2003), among others.
- *Eye tracking* measures eye positions and eye movements to estimate (changes in) gaze direction ([INT 89]). Infrared-based systems exploit the *red-eye effect* that occurs because the cornea and the pupil reflect incident infrared light differently. In contrast, appearance-based systems apply computer vision techniques to the task of first locating the eyes and then determining the orientation of the irises (Jaimes & Sebe 2005; Pantic et al. 2007: 55). Jaimes and Sebe (2005) pointed out that the advantage of infrared-based eye tracking systems over appearance-based trackers is the greater accuracy of the former; however, infrared-based systems may be dangerous for the user's health due to the necessary

prolonged exposure to infrared radiation. Both wearable and non-wearable eye tracking systems have been constructed. The former are more accurate (because head position and direction become available as additional information) but also more intrusive, whereas the latter typically have to be calibrated for each user, which is often a complicated procedure (Jaimes & Sebe 2005).

- *Lip reading* or *speech reading* is the process of observing the visible articulatory movements of a speaker to recognize the words he or she is forming. Recognition may be based on the visual signal alone, or the visual speech information may be used together with the acoustic speech signal, which may be degraded (e.g. in noisy rooms or due to hearing impairment). In computer-based applications, automatic speech reading obtains information about the visual component of the spoken input to improve[146] the performance of automatic speech recognition (cf. Section 11.1.1) in combined *audio-visual automatic speech recognition (AV-ASR) systems* (Potamianos et al. 2003; Potamianos et al. 2009). AV-ASR systems use a visual front-end to extract visual cues of the auditory speech from a video showing the speaker's face (Potamianos et al. 2003: 1307). Three main types of visual speech features have been used in AV-ASR systems (Potamianos et al. 2009). *Appearance-based* features are extracted by analyzing the video pixels in a certain *region of interest (ROI)*, typically the part of the facial image that contains the speaker's mouth region or larger portions of the lower face (Potamianos et al. 2003: 1308f.). The view of *shape-based* feature extraction is that the contours (shape) of the speaker's lips, alone or in combination with facial contours (e.g. of the jaw and cheeks), provide the most relevant information about the visual articulation of speech. This approach has used geometric features, such as mouth height, width, and area, or various parametric models for lip- and face-shape tracking, like the statistical *active shape model (ASM)* (p. 1308). The third approach involves combining appearance- and shape-based features into a joint feature vector that represents both low-level (appearance) and high-level information (shape) about face and lip movements (Potamianos et al. 2009). In general, visual feature extraction requires techniques for locating and tracking the speaker's face, lips, or lower face region that are robust to variations in speakers, pose, lighting, and environment, which still presents considerable problems (Potamianos et al. 2003: 1322). *Audio-visual fusion*, i.e. the integration of the information obtained from the auditory and the visual speech input streams, poses further challenges for research (p. 1307). Current information fusion algorithms can be categorized into *feature fusion* methods, which use a single classifier trained on the combined vector of auditory and visual features; *decision fusion* methods, which recognize audio-visual speech by bringing together the outputs of two modality-specific classifiers for auditory and visual information, respectively; and, finally, *hybrid fusion* methods, which combine characteristics of both feature and decision fusion (p. 1311ff.).
- *Facial expression analysis* (Lisetti & Schiano 2000; Pantic & Rothkrantz 2000; Bartlett et al. 2005; Tian et al. 2005; Cohn & Kanade 2007; Pantic & Bartlett 2007) has largely been based on Ekman and Friesen's *Facial Action Coding System (FACS)*, which defines a set of atomic facial muscular actions called *action units (AUs)* that can be combined to describe different facial expressions (Pantic & Bartlett 2007: 379–383, cf. Chapter 13.4.2). Facial expressions are assigned to different pre-defined categories (often the basic emotions, cf. Chapter 7.15.7

[146] The visual input channel provides support for speaker localization; supplementary information about speech segments; and complementary information about place of articulation for disambiguation (Potamianos et al. 2003: 1306).

and Chapter 10.1.3), depending on the AUs detected in the facial image (Jaimes & Sebe 2005). A number of prototype systems can already recognize a subset of the AUs defined in the FACS from either (near-) frontal views or profile views of a face in an image (sequence) (Pantic et al. 2007: 56). Some approaches are based on detecting and tracking geometric features like facial points or shapes of facial components (e.g. the mouth, the eyes, and the eyebrows). Others are appearance-based, i.e. they apply image filters either to the whole face or to specific regions of the face in order to extract a vector of features corresponding to facial appearance changes that are manifested in the skin texture of the face (e.g. wrinkles, bulges, and furrows) (Tian et al. 2005; Pantic et al. 2007: 56). Still other approaches have used both geometric and appearance-based features, sometimes with better results. Schemes for the classification of facial expressions are static (e.g. Bayesian networks) or dynamic (e.g. hidden Markov models). Static classifiers analyze the information of a single video frame to make a classification, whereas dynamic classifiers consider several video frames, analyzing the temporal patterns of specific facial regions or features. It is easier to train static classifiers, but they also perform less reliably when applied to continuous video sequences. Dynamic classifiers are sensitive to variations in the portrayal of facial expressions across individuals and thus more suitable for person-dependent applications (Jaimes & Sebe 2005).

A major shortcoming of current approaches is that they are more suitable for analyzing elicited, slightly exaggerated expressions of the face under laboratory conditions than for naturally occurring, more subtle facial displays in real-world situations. In response, research has begun on the automatic recognition of AUs in spontaneous facial expressions (Pantic et al. 2007: 56). Furthermore, current techniques commonly assume fairly constrained input conditions and are not robust against variations in face orientation, occlusions, presence of other people, and dynamic backgrounds (Pantic & Bartlett 2007: 408).

- *Gesture recognition* is the process of analyzing human gestures, i.e. volitional or non-volitional movements of different parts of the body that serve a range of communicative functions (cf. Chapter 10.1.2). Gesture recognition requires a mathematical model of the spatial and temporal characteristics of body parts and their movements, techniques for computing the model parameters (representing pose, trajectory, etc.) from the gestural input, and procedures for the classification and interpretation of the gestures portrayed based on the computed model parameters (Jaimes & Sebe 2005). Most gesture recognition techniques fall into one of four categories. *Model-based* approaches build representations of the head, torso, limbs, and fingers from a set of primitive geometric forms including cones, spheres, and cylinders. *Appearance-based* methods track the body and parts of the body using acquired information about color or texture (Pantic et al. 2007: 57). Models using *spatio-temporal salient points* are based on the idea that the content of a scene can be described by acquiring and tracking the rich information in the areas around certain points of interest (e.g. corners and edges) across space and time (Oikonomopoulos et al. 2006). Finally, models based on *spatio-temporal shapes* treat human actions in video sequences as generating shapes in space and time that provide information about the location, orientation, and deformation of the body and its parts at any point in time, as well as dynamic information about bodily movements (Blank et al. 2005). As a whole, the field of human action recognition is still at an early stage. Unconstrained detection and tracking of the hands and the body is not feasible at this point (Pantic et al. 2007: 57), and existing techniques are usually not computationally efficient and highly accurate at the same time (Wang et al. 2003). For a review of vision-based gesture recognition techniques, see the paper by Wu and Huang (1999).

Since Bolt's classic "Put that there" system, which accepted combined inputs consisting of speech and touch-pad pointing gestures (Bolt 1980), more general, more robust, and more natural multimodal interfaces have emerged, in particular in the period since the early 1990s, which has seen considerable progress in the hardware, component technologies, and fusion techniques required for multimodal systems (Oviatt 2003: 287). The majority of multimodal user interfaces that have been developed so far are *bimodal* systems which are capable of processing two parallel input modes, using one or two recognition-based technologies (p. 288). The current state of bimodal interfaces is represented by systems that process either speech and pen-based input (Oviatt & Cohen 2000) or speech and lip movements (Potamianos et al. 2009) in a coordinated fashion (Oviatt 2003: 288). Researchers have also begun to explore the use of vision-based tracking and interpretation of the user's gaze, head position, body location and posture, facial expressions, and manual gestures in multimodal interfaces, which enables monitoring from the background without disturbing the user (Oviatt et al. 2000: 303). However, other sensory modalities besides audition and vision, such as the olfactory and tactile channels, remain largely underexplored in current multimodal systems (Sarter 2006: 441), as are combinations of three or more qualitatively different input modes (Oviatt 2003: 290). In addition, many open research questions remain with respect to both the processing of the individual input modes and their integration. According to a recent article (Reeves et al. 2004), the design of future multimodal interfaces should:

- Accommodate as many and diverse users (cf. Chapter 6.1) and contexts of use (cf. Chapter 6.2) as possible;
- Ensure user privacy and security in the interaction;
- Make the most of each user's cognitive and physical abilities;
- Integrate modalities according to the user's preferences, the context of use, and the desired functionality of the system;
- Minimize user errors and provide effective error handling.

Future multimodal user interfaces will benefit from insights contributed by cognitive science on intersensory perception and intermodal coordination in humans (Oviatt 2003: 293). Another challenge concerns the development of multimodal interfaces with the ability to adapt to the current user, task, dialogue, environmental context, and input modes being used (p. 300). *Adaptiveness* involves the capacity for flexible incorporation and reconfiguration as additional sources of information become available or adverse input conditions or performance failures occur.

11.3 Conversational Agents

Conversational agents were introduced in Chapter 3.2.1 as software agents with the ability to both interpret and respond with spoken or written natural language messages in a contextually and socioculturally appropriate manner in the course of a dialogue with human users, drawing on the human language technologies for speech recognition and synthesis and natural language understanding and generation reviewed in the previous sections. Dialogues with conversational agents are social and/or task-oriented in nature and may integrate other input and output modes besides natural language (cf. Section 11.2), in particular embodied communicative actions (cf. Chapter 10.1) that are interpreted and performed both by the agent and the user. The first (comparatively primitive) instances of these *social agents* (De Angeli et al. 2001b), so-called *chatbots*, are the topic of Section 11.3.2. While chatbots are already widely available in web sites and other applications, most of the more advanced conversational agents and embodied conversational agents still have prototype status. These agents are discussed in Section 11.3.3

and Section 11.3.4, respectively. Section 11.3.1 below summarizes a number of fundamental principles that govern conversations between humans and between humans and agents.

11.3.1 Principles of Conversation

Conversation is a highly contextualized, goal-directed, and cooperative form of linguistic interaction (Levelt 1995: 29f.) involving the coordinated use of language and paralinguistic (non-verbal) cues (facial expressions, gestures, gaze, posture, and locomotion, cf. Chapter 10.1). There are a number of rules and principles that regulate conversations and allow them to be conducted properly (p. 30). First, for the most part, the participants in a conversation do not speak at the same time but adhere to rules of *turn-taking* (cf. Figure 47). A *turn* is a unit within an ongoing conversation that can be attributed to one of its participants and follows and/or is followed by a turn uttered by another participant. Speakers are said to "take turns" during conversation, governed by specific turn-assignment rules which ensure that every participant can contribute to the exchange, that at most one speaker contributes at a time, and that starting, maintaining, and withdrawing from a conversation is performed in an orderly fashion (pp. 31f.). The parties in a conversation normally observe these principles of conduct (p. 37), but there are also often instances where turns overlap, one speaker dominates the exchange, communication breaks down, and so on.

During conversations, interlocutors commonly make reference to ("point at") people, places, points or periods of time, and elements of the discourse, from their own perspective and relative to the current context, by means of *deixis*, which, in face-to-face conversations, involves the use of natural and unambiguous combinations of speech, gestures, and locomotion (Lester et al. 1999b). Four types of deixis are commonly distinguished (Levinson 1983: 62f.; Levelt 1995: 44–58; Cruse 2000: 319–324):

- *Person deixis*. Realized in language by personal pronouns, this type of deixis enables the speaker to refer to himself or herself (the first person), the addressee(s) (the second person), and other individuals and objects (the third person) that are neither speaker nor addressee.
- *Place deixis*. Speakers use expressions that locate people and objects in space relative to the parties in the conversation, in order to identify those entities through their location, inform addressees about their whereabouts, or acknowledge the location of an entity.
- *Time deixis*. Deictic reference to time locates points and spans on the time axis relative to the time of speaking. Speakers may refer to times before, at, or after the time of their utterance.
- *Discourse deixis*. Utterances in an ongoing conversation may include deictic expressions that relate them to elements of the previous, current or future discourse, which facilitates hearers' orientation within the exchange.

Conversations are generally characterized by a tacit mutual assumption of *cooperativeness* among their participants. In other words, the parties in a conversation usually expect that the contributions of other interlocutors will pertain to their own utterances and to the overall goal of the exchange (Cruse 2000: 355). All contributions are implicitly assumed to be governed by a *cooperative principle*, which has been formulated as follows:

> *Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (Grice 1975: 45).*

What it means to be 'cooperative' in conversation according to this general principle is spelled out by four more specific subprinciples, the so-called *conversational maxims*, which

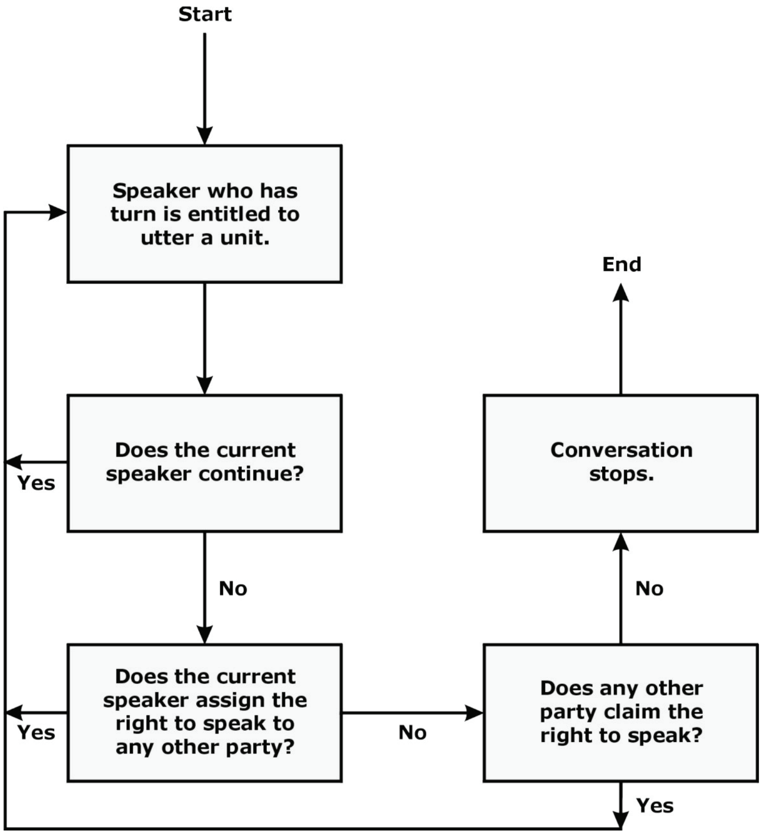

Figure 47. A model of turn-taking. Adapted from ([INT 90]).

cooperative speakers follow in designing their contributions (Grice 1975; Levinson 1983: 101f.; Cruse 2000: 355ff.):

- *The maxim of quantity*. Be sufficiently informative. Do not provide more or less information than is required.
- *The maxim of quality*. Be truthful. Do not make knowingly false or unsupported statements.
- *The maxim of relation*. Be relevant. Do not say things that are unrelated to the current conversation.
- *The maxim of manner*. Be perspicuous. Avoid obscurity and ambiguity (in context). Be brief (not wordy) and orderly (i.e. relate events in the order of their occurrence).

These maxims are not strict rules in the sense of rules of grammar (cf. Section 11.1.3.2.1). They are better thought of as flexible guidelines (Cruse 2000: 357), which may be violated or flouted by speakers. When speakers *violate* maxims, the lies, omissions, ambiguity, irrelevance, or obscurity in their utterances will not be obvious to hearers. As a result, communication might be impaired. In contrast, a hearer will notice when a speaker *flouts*, i.e. deliberately and openly departs from, one or several maxims in his or her utterance to achieve a certain communicative effect (Peccei 1999: 27f.; Cruse 2000: 360). Based on the assumption that the speaker adheres to the cooperative principle, the hearer will try to infer what the speaker attempts to communicate beyond the semantic content of his or her contribution. The resulting inferences, which allow the hearer to relate the speaker's utterance to the ongoing conversation in a meaningful way, are referred to as *conversational implicatures*. For example, if a lecturer asks a college student "Why didn't you do your homework?" and the student replies "There was a party in our dorm," the student flouts the maxim of relation. However, the lecturer can still relate the response to his or her question on the basis of the implicature that participating in the party (and getting drunk) kept the student from doing his or her homework.

While the cooperative principle offers an explanation of how implicatures arise, it fails to account for why people commonly choose to convey what they mean in an indirect way, in other words, why speakers are normally polite (Cruse 2000: 361). *Politeness* is instrumental in staying on good terms with others, even when it is necessary to communicate things that may be unpleasant for the hearer (p. 362). In particular, politeness plays an important role in helping to preserve the hearer's public self-image (his or her social 'face') (Goffman 1955; Goffman 1967, cf. Chapter 7.14). The notion of *face* refers to the mask created and maintained by people to present themselves to others in social situations. People wear different masks for different audiences and social interactions (Goffman 1967). Face can be subdivided into two categories: *positive face* is the desire to have the appreciation, approval, and respect of others, whereas *negative face* is the want for autonomy, i.e. for freedom from imposition or impediment from others (Goffman 1967; Brown & Levinson 1987). In conversation, certain communicative acts, such as criticism, suggestions, and requests, can threaten the positive and/or negative face of the other party, although they may be necessary to achieve the speaker's communicative goals (Mayer et al. 2006: 36; Wang et al. 2008: 99). Acts of a speaker that may pose a threat to the addressee's face are called *face-threatening acts (FTAs)*. For example, criticism may threaten a person's positive face because it suggests that the speaker does not appreciate or approve of the addressee. Threats to negative face can occur when, for example, advice is perceived as impeding the addressee's freedom to decide for himself or herself (Mayer et al. 2006: 36f.). In Brown and Levinson's theory of politeness (Brown & Levinson 1987), model speakers modify their language to design messages that attempt to maintain the face of others while achieving the speaker's goals at the same time. *Politeness* thus signals that the speaker intends to reduce threats to the hearer's face posed by particular face threatening acts (Mills 2003: 6). Brown and Levinson (1987) described four politeness strategies for handling potential face threats in conversation:

- *Bald on record*. Make no attempt to mitigate threats to the addressee's face. Use direct language; do not hedge opinions, minimize imposition, or attempt to avoid disagreement.
- *Positive politeness*. Minimize threats to positive face by respecting the addressee's want for the appreciation and understanding of others. Strengthen the effect of commands, offers, requests, etc. that involve a benefit to the hearer by making them more direct or less optional (Cruse 2000: 363f.).
- *Negative politeness*. Minimize threats to negative face by honoring the addressee's wish to maintain his or her autonomy. Weaken the effect of commands, offers, requests, etc. that involve a cost to the hearer, by making them more indirect or optional (Cruse 2000: 363f.).
- *Off-record (indirect) strategies*. Do not impose directly on the addressee. Use indirect language (e.g. hints, vague statements, sarcasm, etc.) to insinuate your intent without directly stating it.

The *politeness principle* requires speakers to phrase their utterances in a way that minimizes the expression of impolite beliefs (Leech 1983: 81), i.e. disparages the hearer's status as little as possible (Cruse 2000: 362). If possible, they should avoid bossing, slander, Schadenfreude, disagreement, and self-praise when they talk to the hearer (p. 362).

In the absence of cues to the contrary, the participants in a conversation implicitly assume mutual cooperativeness and politeness by default. In other words, the general tacit agreement between conversation partners is that they will make an effort to understand each other and to avoid hurting each other's feelings (Nass & Brave 2005: 182). This effort involves speaking intelligibly, being an attentive and sensitive listener, adhering to the mechanisms of turn-taking, and respecting both the cooperative and the politeness principle when formulating responses to the contributions, thoughts, and feelings of the other participants (p. 182). But communication can still fail at some point during a conversation. In this case, cooperation either may come to an end or it can be revived by a joint effort of the participants who hold on to the mutual assumption that the other's contribution was cooperative (p. 182).

The conversation principles discussed in the previous paragraphs have often been neglected in the design of dialogues involving the computer as an interaction partner because knowledge of proper conduct in conversations is tacit rather than explicit and thus tends to escape the attention of conversational (in particular voice) interface designers (Nass & Brave 2005: 179). Hence, it is not surprising that the resulting interfaces, while using language, are not perceived as overly conversational and often leave the user frustrated and the dialogue unfinished (p. 179). Nass and Brave discussed three conversational maxims that many current voice interfaces fail to observe: quantity, relation, and manner (pp. 179–182). Unnecessarily verbose or insufficiently detailed prompts as well as overly general questions (e.g. "How can I help you?"[147]) are examples of violations of the maxim of quantity. Voice interfaces depart from the maxim of relation when they ask for information (e.g. about the user's credit card) before the user deems it relevant for advancing the conversation. Finally, they become hard to understand (thus violating the maxim of manner) when they assume too little or too much background knowledge of users and the complexity of their messages is therefore not appropriate for the current user.

Viewing human-agent dialogues as respectful cooperation and designing conversational agents accordingly to make meaningful use of turn-taking, deictic behaviors, and cooperative and polite language can facilitate the process of achieving mutual understanding in interactions between agents and users. Humans have a preference for cooperativeness and politeness in their interlocutors, and they are likely to look for these qualities in conversational agents as well.

11.3.2 Chatbots

The first generation of conversational agents is beginning to spread throughout the World Wide Web as *chatbots*,[148] software systems that simulate conversation, or chat, with human visitors in chat rooms or on web sites (De Angeli 2005). Chatbots are designed to imitate human conversational behavior as closely as possible, often in order to induce the belief in users that they are talking to another human being and in this way to pass the *Turing Test*. This is an empirical test proposed by the British mathematician Alan Turing (1912–1954), in which a machine's ability to think is assessed based on its use of language (Jurafsky & Martin 2000: 7). The test is an imitation game with three players, two of which are human while the third is a computer. One of the two people and the computer are hidden from the view of the second

[147] Unless a conversational interface can handle unrestricted user input (which is beyond the capabilities of any system currently available), this question does not provide enough information for users to understand what inputs the interface can handle. Nass and Brave suggested using a more specific question which makes the options (exemplified by A, B, and C) available to the user explicit: "Do you need help with A, B, or C?" (Nass & Brave 2005: 181).

[148] The term 'chatbot' is a shortened form of "chat robot." Alternative labels for these conversational programs include 'chatterbot' or just 'bot.' The term 'chatterbot' first appeared in a 1994 conference paper by Michael Mauldin (Mauldin 1994).

human who interrogates them via a teletype[149] to determine from their written replies which one is the computer. The goal of the machine is to make the interrogator think that it is the human by answering his or her questions like a real person would (to the best of its abilities), whereas the hidden human being attempts to convince the questioner that he or she is human and that the other player is the machine (p. 7). While the Turing Test as an empirical method to assess machine intelligence was proposed already in 1950 (Turing 1950), it still inspires debate and experiment (Copeland 2000; Saygin et al. 2000) more than fifty years later. Turing himself expressed confidence that machines would be able to think one day (Turing 1950).

Based on the absence or presence of deception in the interaction between a user and a chatbot and the user's (un-) awareness of it, De Angeli (2005) proposed a taxonomy of chatbots that is shown in Table 16. *Explicit chatbots* make no pretence of their status as artificial entities, portraying themselves in a way that helps the user to understand that he or she is talking to a machine, which may include explicit instructions for conversation. These bots do not try to deceive but rather cooperate with the user in building a virtual persona. In contrast, *deceptive chatbots* are designed to fool unaware users into accepting them as humans. A famous example from the early 1990s is Julia (Foner 1993; Mauldin 1994; Turkle 1995: Chapter 3; Murray 1998: Chapter 8), a chatbot that lives on MUDs,[150] chats with users, answers questions about the MUD (and other topics), plays the card game Hearts, keeps track of other users, relays messages, remembers things, sings songs, and gossips (Murray 1998: 215f.). As a female impersonator, Julia was a success. It has been reported that one user even tried to seduce the bot into a sexual relationship (pp. 216f.). Finally, *competitive chatbots* make the explicit claim to be human and invite their human interaction partners to prove them liars, usually in the context of some online community or competition dedicated to the purpose, such as the Loebner Prize ([INT 91]), an annual contest since 1990 in which several candidates compete for the title of "most human-like chatbot" in a Turing Test setup. When talking to a competitive chatbot, the user is aware that his or her conversation partner is a computer program which will do its best to substantiate the claim that it is human. The user's task is to trick the chatbot into betraying itself by eliciting a response during the conversation showing that the bot is not human as it claims to be.

De Angeli pointed out that there are no fixed boundaries between these categories. The same chatbot may embody different types in different contexts of interaction. It often happens that signals from chatbots are inconsistent (cf. Chapter 7.5). At times, they may pretend to be human and at others to be an artificial creature, or they may show a photorealistic human embodiment (cf. Chapter 9.5) while telling the user that they are machines. The user's awareness of the status of a chatbot can also shift during an interaction (usually because he or she realizes the bot's deception) (De Angeli 2005).

One of the oldest and best-known chatbots is ELIZA (Weizenbaum 1966), which showed quite impressively how a computer program can be designed to be *perceived* as understanding while in fact it does not understand at all. Playing the role of a Rogerian psychotherapist, in

[149] A *teletype* is a typewriter-like device for sending and receiving typed text messages through an electric communication channel. Teletypes, or *teleprinters*, as they are also called, were widely used for the larger part of the 20th century but are nowadays obsolete ([INT 92]).

[150] A *MUD (multi-user domain/dungeon/dimension)* is a multiplayer networked computer game environment created and maintained through textual descriptions of rooms, objects, events, player and non-player characters as well as the interaction of distributed players with each other and their surroundings using text messages and commands sent over the network (usually the Internet) (Bartle 1990; Bartle 2003; [INT 93]).

Table 16. A taxonomy of chatbots. Adapted from De Angeli (2005, Table 1).

Chatbot Type	Deception	User
Explicit	Absent	Aware
Deceptive	Present	Unaware
Competitive	Present	Aware

conversations with the user, ELIZA applied pattern matching techniques to produce responses that often involved the transformation of an input statement into a question to which the user responded with another statement, thus restarting the cycle.[151] Most current chatbots ([INT 94]; [INT 95]; [INT 96]) are essentially only slightly improved variants of ELIZA's original approach (De Angeli et al. 2001a).

Figure 48 shows the architecture of a chatbot. While the underlying algorithms of chatbots vary in complexity (De Angeli 2005), most bot programs associate a *stimulus* (i.e. a user input) with a *response* (i.e. a pre-scripted statement, or *template*, which may have both variable and functional parts) (e.g. Wallace 2008). Stimuli typically consist of words and sentences typed by the user; responses are delivered as text ([INT 97]).

After an initial stage performing spelling correction, input normalization, and other pre-processing operations, an efficient search algorithm is used to compare the pre-processed input against a database of *response categories* (hierarchically nested pattern-template (i.e. question-answer) pairs) (Wallace 2008: 193f.) in order to find a matching pattern (a keyword, phrase, or whole sentence) ([INT 97]). The template associated with the found pattern is returned as the chatbot's response, possibly after filling in missing information with values from a model of the user (e.g. the user's name), or updating the user model with information derived from the input (e.g. sex, age, favorite color, etc.). More sophisticated chatbots have the ability to enhance their response capabilities by automatically detecting new input patterns in log files of user-bot conversations and having a human developer author appropriate response templates (Wallace 2008: 182). Schumaker et al. (2006) described a chatbot system that included a module which allowed users to rate the bot's responses and suggest corrections.

In web sites, chatbots are often portrayed as animated characters with at least a partial body (cf. Chapter 9.6), which allows them to accompany their messages with rudimentary non-verbal communicative (cf. Chapter 10.1) and believability-enhancing (cf. Chapter 10.2) behaviors. Today, many interactions with chatbots are still text-based (Coniam 2008: 100), in particular on the World Wide Web, where the feasibility of other input modes apart from keyboard and mouse input is still limited (cf. Chapter 7.10). However, speech recognition (cf. Section 11.1.1) and speech synthesis (cf. Section 11.1.2) are beginning to be used in chatbots (Atwell 2005; Coniam 2008: 100).

Chatbots are the first step toward achieving the old dream of building technology-based artificial companions for people (Bledsoe 1986; De Angeli et al. 2001a). While chatbots are often used to provide some kind of service to users, their major purpose is social rather than

[151] Despite its simplicity, ELIZA was surprisingly successful with users. Some of them responded to the program in the same way as they would to a human therapist (Fogg 2003: 111). The implications of these reactions for human-computer relationships deeply troubled Joseph Weizenbaum (1923–2008), ELIZA's creator. In an influential book (Weizenbaum 1976), he argued that there should be limits on what humans allow computers to do because even a machine with artificial intelligence will never have compassion, wisdom, and other qualities that humans possess by virtue of having been raised in a family ([INT 98]).

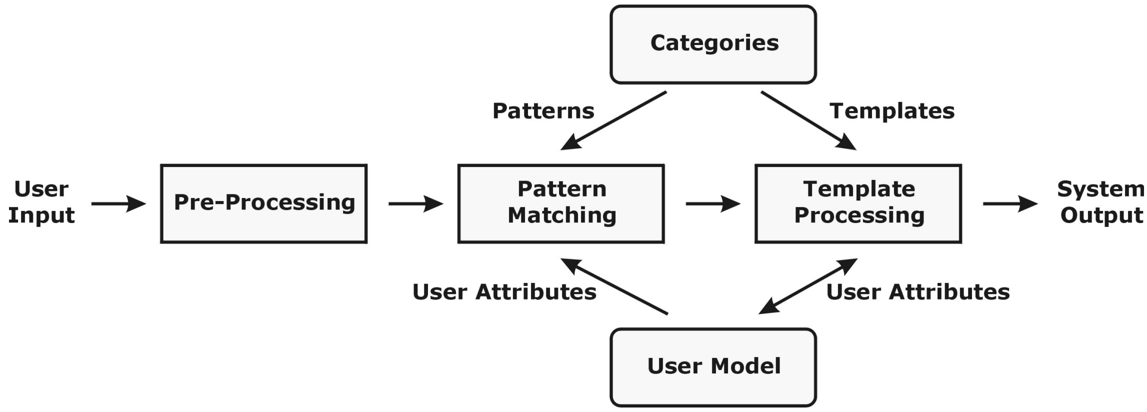

Figure 48. The architecture of a chatbot.

task-related in nature (De Angeli et al. 2001b). Being *social agents* (De Angeli et al. 2001b), chatbots need to satisfy requirements beyond usability (cf. Chapter 6.3). In particular, they have to be designed to facilitate establishing and maintaining relationships with users (cf. Chapter 7.16). On the World Wide Web, these social agents can be constantly available in chat rooms or on web sites to answer the questions of users and keep them company. Since both profit and non-profit organizations are under more pressure than ever to deal more effectively with users' inquiries and to ensure their loyalty (De Angeli et al. 2001a; De Angeli et al. 2001b), these features of chatbots invite their use as virtual assistants, guides, salespeople, and entertainers in order to form long-term social relationships with users (Morel 2004: 179ff.; see Abu Shawar & Atwell 2007 for a discussion of applications of chatbots). In particular, the hope is to achieve cost reductions by using chatbots to automate routine tasks (e.g. providing information about products, responding to frequently-asked questions), and to increase brand or product awareness and customer loyalty by creating a more enjoyable experience for the user. Finally, chatbots can elicit valuable personal information from their human interaction partners by casually asking questions during a conversation (De Angeli et al. 2001a).

To serve these functions well, chatbots must be competent, eloquent, and interesting to users. However, developing or customizing even a moderately sophisticated chatbot can be quite expensive as its behavior and identity depend on its database of response categories, and building or changing this database may require considerable effort. The response categories play such a central role because a chatbot essentially defines itself through its responses, revealing who it is, what it thinks, and how it feels by way of its reactions to user inputs.[152] Therefore, the chatbot's responses should adhere to the design principles of consistency, coherence, individuality, and variability introduced in Chapter 7. For embodied chatbots, the body and its non-verbal behaviors are additional design elements to be considered, which may complement or clash with the chatbot's verbally created persona, so there is also the need to reconcile linguistic responses with the appearance and behaviors of the chatbot's body. As the

[152] Russell (2002) described an experiment involving a chatbot that portrayed two different personality types (extrovert and introvert, cf. Chapter 13.5) through linguistic properties of its responses, including style (utterance length, use of tag questions, and hedging) and lexical selection (number of intensifiers and emotion words) (cf. Section 11.1.4). Results indicated that "personality effects, in such an interface, are more complex and subtle than previous research has suggested and that to produce 'real' synthetic personalities, the interaction style as a whole must be considered" ([INT 99]).

set of response categories of a state-of-the-art chatbot is typically poorly structured, consisting of a collection of large text files, it becomes difficult to stay on top of all the aspects of the content and wording that contribute to the consistency, coherence, individuality, and variability of responses.

Practitioners suggest that creating a chatbot for a business or other organization should proceed in analogy to recruiting a new employee through a series of three stages (Morel 2004: 191):

- *Head hunting*. Identify objective and subjective aspects making up the character profile (cf. Chapter 9.3) of the chatbot to be recruited. Determine the bot's function, required expertise, and future workplace. Design a chatbot that meets these requirements.
- *Training*. Teach the chatbot how to play its role competently and convincingly for the target group of users, which especially includes giving appropriate responses to user inputs in one or several languages.
- *Contract*. Deploy the new chatbot in its work environment. Make the necessary adjustments and improvements. The bot may have a short-term (e.g. for a special campaign) or a long-term contract (e.g. a permanent position in customer support) with its employer.

Carefully designed chatbots can successfully create (at least temporarily) the illusion of intelligence and make the user suspend his or her disbelief (De Angeli et al. 2001a). While some users quickly identify the limitations of a chatbot, others will continue to chat with the bot for hours without even noticing that they are talking to a machine. Overall, chatbots are a viable option for designers of interactive systems if the application involves answering simple questions or engaging the user in small talk (cf. Chapter 10.3). Coniam reviewed five current chatbots and concluded:

> *[T]he chatbots examined are still a long way from passing the Turing Test. The current generation of chatbots cope best when presented with one-clause questions or statements embracing straightforward propositions with minimal cohesive linkage to previous utterances. (Coniam 2008: 115).*

However, today's chatbots are less suitable for more complex tasks, such as counseling or tutoring, because they lack everything beyond the most basic dialogue processing capabilities. A chatbot is generally unable to carry on a structured, goal-directed dialogue with the user. More often than not, after only a few turns, it becomes obvious that the conversation is going nowhere. As a result, chatbots are not appropriate for educational applications in particular, given that dialogues with learners should be both structured and directed toward a well-defined instructional goal (cf. Chapter 12.3.9). Usually, they also tend to last longer than a few turns. Conversational agents for education need more sophisticated dialogue capabilities than current chatbot engines can offer (Coniam 2008). As a side note, developers of (embodied) conversational agents have used chatbot engines, despite their limitations, to implement the dialogue component of their agents (e.g. De Angeli et al. 2001a; De Angeli et al. 2001b; Tatai et al. 2003; Abbattista et al. 2004; Laufer & Tatai 2004; De Angeli 2005; Goh et al. 2006), mainly for two reasons: (1) a large number of chatbots with quite sophisticated category databases are freely available (e.g. [INT 94]); (2) chatbots provide a quick off-the-shelf solution for dialogue processing in agents when this capability is required but not a focus of the research. However, such compromises tend to lead to an imbalanced design (cf. Chapter 7.3).

11.3.3 Dialogue Agents

Conversational agents using more sophisticated speech and language processing than chatbots to enable more flexible and purposeful interactions are increasingly finding their way from research laboratories into practical applications (McTear 2004). These *dialogue agents* include more advanced technologies for interpretation and response generation and, importantly, feature a central component for managing the dialogue with the user. The diagram in Figure 49 shows the architecture of a dialogue agent, which consists of three major processing stages (Lester et al. 2004: 10-4ff.):

- The *interpreter* identifies and interprets the sequence of words provided by the user and computes a representation of the meaning of the input.
- The *dialogue manager* processes the meaning representation and decides on an appropriate reaction, interacting with data stores and external applications.
- The *response generator* produces the response and delivers the output.

Four major resources support dialogue processing at all stages: the dialogue strategy, the dialogue model, the user model, and mechanisms for error handling. The processing stages and resources of a dialogue agent are discussed in the following sections.

11.3.3.1 Interpreter

The first processing step for a dialogue agent is to *interpret* the user's utterance. This requires natural language understanding technology (cf. Section 11.1.3), which transforms the input string of words into a meaning representation for use by the dialogue manager (McTear 2002: 103). If the agent accepts spoken input, NLU has to interact closely with a speech recognition component (cf. Section 11.1.1), which maps from the user's spoken utterance to a string of words. A simple way to integrate speech recognition and natural language understanding in spoken dialogue agents is to sequence two independent systems, with the recognition module sending the word-string with the best recognition score to language understanding (Price 1997: 45; McTear 2002: 109). Since this strategy can easily miss the actual sequence of words, alternative approaches to the integration of speech recognition and natural language understanding have been used (Holmes & Holmes 2001: 204f.; Huang et al. 2001: 664–670; McTear 2002: 109f.; Jurafsky & Martin 2008: 337f.):

- *Word lattice parsing*. The speech recognizer outputs a network of possible word sequences (called a "word lattice" by convention) that can be visualized as a directed graph in which nodes represent points in time and arcs represent scored word hypotheses. The NLU module then tries to determine the best-scoring grammatical utterance that covers the entire spoken input.
- *N-best filtering*. In this approach, the speech recognizer sends a rank-ordered list of its n-best (n = 10 .. 100) sentence hypotheses to language understanding. The NLU module then uses more sophisticated knowledge sources to re-score and re-rank the n-best hypotheses in order to find the "1-best utterance" with the highest score.

Traditional NLU techniques cannot be easily used to interpret spoken language because they have been designed to process *text* input consisting of grammatical sentences of unmistakably identifiable words. In contrast, an agent for spoken conversation has to deal with unknown

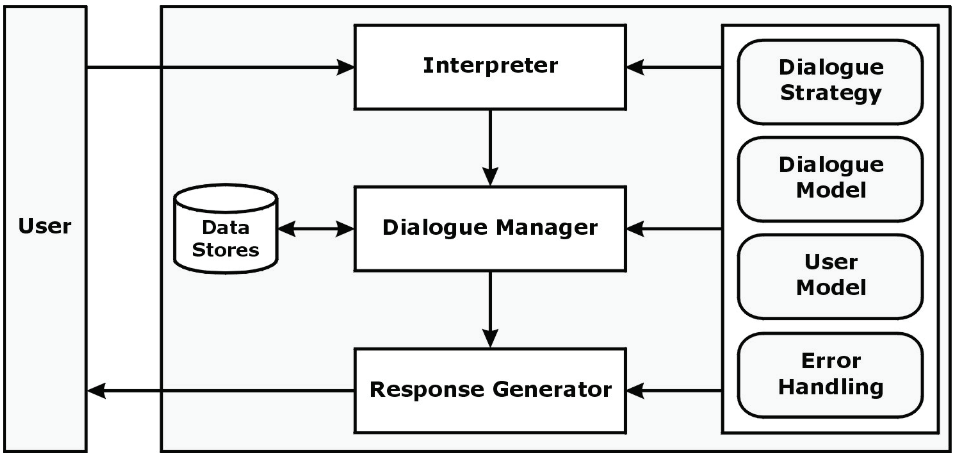

Figure 49. The architecture of a dialogue agent. Based on Aust and Schröer (1998, Figure 1) and Lester et al. (2004: 10-5, Figure 10.2).

words and constructions, recognition errors, and, most importantly, instances of spontaneous speech (Zue 1997).

Spontaneous speech consists of utterances created by speakers "ad lib" (rather than based on careful advance preparation) in a particular context as they are trying to communicate messages to others. Spontaneous utterances are commonplace in everyday conversation, whether face-to-face or over the telephone. Their generation is subject to constraints imposed by the limited time and cognitive resources available to the speaker (Androutsopoulos & Aretoulaki 2003: 636). As a result, they contain many *disfluencies*[153] (Brennan & Schober 2001; Eklund 2004). Examples of these disruptions to the normal flow of utterances include filled pauses (e.g. *Erm*, *Eh*[154]), full or partial word repetitions (e.g. *NoNo*, *aFlightaFlight*), false starts (e.g. *WhatD'You-I'dLike*) and self-repairs (e.g. *FromGa-FromLondonGatwick*), coughs, tongue and lip clicks, etc. Ellipsis (i.e. omission of sentence parts, e.g. *NotAachenAthens*) and anaphora (i.e. back references by means of e.g. pronouns) are further common elements of spontaneous utterances. The input speech signal may also be distorted by background noise and channel distortions (e.g. over the telephone) (Androutsopoulos & Aretoulaki 2003: 635f.). Previous research has tended to focus largely on read speech and written text; hence, understanding of spontaneous speech phenomena is still limited.

Dialogue agents that are confronted with spontaneous user utterances have to work with the (most likely incomplete and/or incorrect) sequence of words that they were able to recognize in the continuous stream of input speech. Since it would be futile to attempt an in-depth analysis of every utterance, in practice, *robust parsing* techniques (Jurafsky & Martin 2008: 450–456, Chapter 14) are employed to extract all useful bits and pieces of information from the input while disregarding the rest. In particular for elliptical or ungrammatical input utterances, this involves parsing constituents (i.e. words and phrases) by themselves rather than trying to integrate them into an overall sentence structure (Androutsopoulos & Aretoulaki 2003: 637f.). Approaches to robust parsing make use of probabilistic context-free grammars, semantic grammars, and shallow parsing (cf. Section 11.1.3).

[153] An in-depth account of disfluency from a wide range of perspectives can be found in Eklund (2004).
[154] All examples were adapted from Androutsopoulos and Aretoulaki (2003: 635f.).

11.3.3.2 Dialogue Manager

The ability of dialogue agents to effectively manage their exchanges with users is the key to successful human-agent conversations. If the dialogue is conducted carefully, user satisfaction increases (thanks to a larger number of successfully completed interactions), and it becomes possible to compensate for shortcomings of the human language technologies involved (cf. Section 11.1), in particular recognition and understanding errors (Mangold 1994: 98). Hence, dialogue management is generally regarded as an enabling technology for dialogue agents (e.g. Cole et al. 1995).

Dialogue agents typically have a central module, the *dialogue manager*, which controls the interaction with the user and coordinates the other system components (McTear 2002: 103) to ensure that the conversation flows effectively and naturally over several dialogue turns between human and agent toward a valid and satisfactory conclusion (Giachin 1997: 210). This involves the following tasks:

- Detecting errors in recognition and understanding and recovering from them;
- Preventing conversational breakdowns and failure;
- Interpreting user utterances in the context of the dialogue and task;
- Initiating appropriate clarification and confirmation subdialogues if necessary;
- Predicting what the user can say in the next dialogue turn;
- Interacting with the application back-end in service of the user's needs;
- Determining the relevant information to be conveyed to the user;
- Coordinating understanding and response generation components.

To fulfill these tasks, the dialogue manager interacts with four major resources, including a strategy driving the conversation, an evolving model of the dialogue, a set of assumptions about the user, and strategies for detecting and handling errors in the interaction. These resources are discussed below.

11.3.3.2.1 Dialogue Strategy

Dialogue agents should manage conversations with the user according to a *dialogue strategy* that steers the conversation (Giachin 1997: 211) and determines the allocation of the *initiative* (i.e. which of the participants is in control of the exchange at any point in time during the dialogue). Dialogue strategies differ mainly with respect to the degree of user control vs. system control in the conversation (Huang et al. 2001: 860; Androutsopoulos & Aretoulaki 2003: 641f.; McTear 2004: 108ff.; Jurafsky & Martin 2008: 828f.):

- *System initiative*. The agent is completely in charge of the exchange. It asks the user a series of questions in order to elicit the information necessary to complete the task, and expects the user to only answer the current question at every step of the dialogue. The possible dialogue states and conversation paths are pre-defined in a network (cf. Figure 50). Each dialogue state consists of a prompt and an associated set of allowed user responses. On the one hand, constraining user inputs this way facilitates dialogue processing considerably. On the other hand, the inflexibility of such structured human-machine dialogues may quickly lead to user dissatisfaction with the interaction and hence the agent.
- *User initiative*. The agent is responsive but never proactive in the course of the dialogue (cf. Chapter 3.1), i.e. it only responds to the inquiries of the user, who controls the exchange.

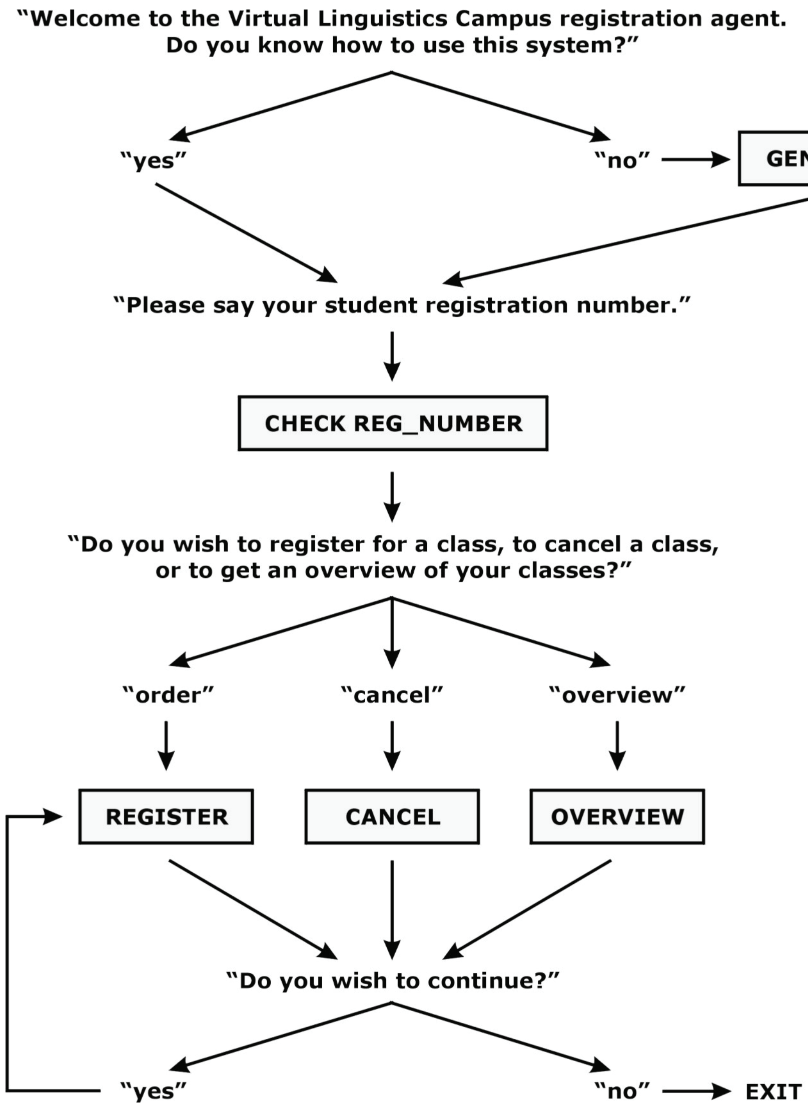

Figure 50. Simple dialogue-state network for a conversational registration agent adopting a system-initiative dialogue strategy. Adapted from McTear (2002: 129, Figure 11).

This strategy is typical for agents that respond to users' requests for information or carry out their commands. While letting the user lead the dialogue is an attractive option in general, it becomes difficult to successfully complete the interaction in the event of communication problems or missing information since the agent cannot take control to initiate a subdialogue that clarifies misunderstandings or asks for the required items. Furthermore, users may feel uncertain about the agent's exact range of capabilities and become frustrated when their expectations are not met (cf. Chapter 7.3).

- *Mixed initiative*. Mixed or *variable initiative* dialogues are characterized by shared control between the user and the agent. In the course of a dialogue, the initiative may shift from the one to the other (and back again) according to the requirements of the current state of the interaction. The user remains in control for as long as he or she does not seem confused and the agent is able to determine his or her intention. However, the agent is always ready to guide the user or engage him or her in a clarification subdialogue if necessary.

In addition to strategies for handling the dialogue with the user, dialogue agents also require ways to deal with situations where they detect a potential misrecognition or misinterpretation of the user's utterance and need to verify, or confirm, by consulting the user that they understood him or her correctly (Androutsopoulos & Aretoulaki 2003: 642f.; McTear 2002: 117ff.; McTear 2004: 118–121; Jurafsky & Martin 2008: 831f.). *Explicit verification* involves asking the user directly for confirmation of a particular piece of information acquired, typically in the form of a yes-no question to restrict what the user can say in response (Jurafsky & Martin 2008: 831). This is a robust verification strategy, but the dialogue becomes longer and more awkward, which has a negative effect on user satisfaction (McTear 2002: 117; Jurafsky & Martin 2008: 831). *Implicit verification* is an alternative approach in which the agent phrases its next question in a way that repeats the agent's understanding of the user's previous utterance(s). The repeated information is implicitly confirmed if the user's follow-up response contains no corrections (McTear 2004: 120). This strategy is more natural than explicit verification, but it presupposes a certain level of confidence in the results of the recognition and understanding stages (McTear 2002: 117f.). Furthermore, the range of possible user utterances in response to an implicit verification request is larger, which increases the complexity of the interpreter module (p. 118). In addition, such requests combine old items to be confirmed with new items to be elicited. The number of possible combinations can quickly grow to such an extent that dynamic generation of the verification questions (cf. Section 11.1.4) is the only feasible option (p. 118).

11.3.3.2.2 Dialogue Model

The *dialogue model* provides a continuously updated representation of the structure of the ongoing agent-user dialogue, which contains information about the dialogue history, indicates the current *dialogue state* (i.e. the state of the interaction with respect to the goal of the exchange), and allows predictions about future interactions (i.e. about what both the user and the agent can possibly say in the next dialogue turn) (Cole et al. 1995; Cohen 1997; McTear 2002).

The *dialogue history* keeps track of what has happened before in a given dialogue, indicating the objects, actors, and events mentioned by the participants and providing information about their relationships to one another (McTear 2002: 95). This information is useful for identifying the referents of anaphoric expressions (including pronouns, possessive determiners, and definite noun phrases) (p. 95). Furthermore, for every user utterance and agent response, the dialogue history may indicate the *dialogue act*[155] performed plus associated task parameters and their values. Using this information, elliptical user utterances can be interpreted by anchoring them in the context of the previous exchange (Androutsopoulos & Aretoulaki 2003: 639).

Agents that model the evolving structure of dialogues with the user can act more flexibly and robustly in the course of an interaction. Cole et al. (1995) identified several benefits of dialogue modeling. First, access to the previous context makes it easier for the agent to recover from recognition and understanding errors. Second, expectations about the contents of incoming utterances based on the agent's model can even help to avoid such errors in the first place. Third, dialogue modeling helps the agent to keep the conversation on track towards the goal despite more or less severe errors. Fourth, the interaction becomes more efficient because

[155] A *dialogue act* is a kind of speech act performed in a contribution to a dialogue in order to further a particular (sub-) goal of the exchange. See Jurafsky and Martin (2000: 727–744; 2008: 840–846) for a discussion of dialogue acts and their automatic interpretation.

shorter and simpler user utterances and fewer dialogue turns are required. The agent's outputs can be tailored to present only information that is important for the conversation to progress effectively.

Three major approaches have been used to model dialogue structure (Jurafsky & Martin 2000: 750–757; Huang et al. 2001: 886–894; McTear 2002: 92ff.; Leech & Weisser 2003: 152f.; McTear 2004: 111–117; Jurafsky & Martin 2008: 827–831). The first, *finite-state based dialogue modeling* has been used in many (in particular commercial) systems (McTear 2002: 92; Androutsopoulos & Aretoulaki 2003: 639; Jurafsky & Martin 2008: 827). This method makes use of *dialogue grammars* which impose a fixed structure on dialogue that is represented as a finite-state transition network (cf. Section 11.1.3.2.1) whose states correspond to dialogue acts, such as `greeting`, `request_date`, and `list_flights` (cf. Figure 51) (Leech & Weisser 2003: 153). The network is organized based on the concept of *adjacency pairs* (e.g. question–answer, greeting–greeting, challenge–rejection, etc.) (Huang et al. 2001: 887). Transitions between states occur on the basis of tests that check if the representation derived from the user's utterance contains the required information. The system controls the interaction, prompting the user to provide information for the current dialogue state and proceeding to another state based on the user's response (McTear 2002: 93). Finite-state models continue to be popular because of their simplicity. They are suitable for modeling structured task-oriented dialogues directed by the agent (system initiative, cf. Section 11.3.3.2.1) that involve eliciting information from the user in a pre-defined sequence (McTear 2004: 111f.). However, they are too restricted to be able to support flexible and natural human-machine dialogues that allow freedom of expression, shift of initiative, and deviations from pre-specified dialogue paths (McTear 2002: 128; McTear 2004: 112).

Frame-based dialogue agents complete (or fill in) a pre-set frame (or *template*) with pieces of information obtained from the user during the conversation (McTear 2002: 93f. + 131–134; McTear 2004: 113–116; Jurafsky & Martin 2008: 829f.). This frame functions as a dialogue model that keeps track of missing items (McTear 2002: 131). Properties of interactions with a frame-based agent include the following (McTear 2002: 93f. + 131–134; Jurafsky & Martin 2008: 829):

- Information does not have to be elicited in a pre-determined order.
- Mixed initiative is possible. The dialogue is driven by the information provided by the user and required by the agent.
- The agent can integrate pieces of information that were supplied by the user before the agent explicitly asked for them.
- The agent does not ask more questions than necessary because it knows what values are still required to fill the frame.
- Questions and other prompts will only be produced if their pre-conditions are fulfilled.
- The agent's input analysis focuses on extracting task-relevant information from the user's utterance (shallow parsing, cf. Section 11.1.3).

The usefulness of frame-based dialogue models is limited to well-defined tasks involving a homogeneous range of users, a static world, and a collection of recurring, precisely specified operations (e.g. database queries) (McTear 2002: 134).

The basic idea of *plan-based dialogue modeling* (Allen 1994: 480–483; Huang et al. 2001: 888–892; Jurafsky & Martin 2008: 850–855) is that communicative actions of (both human and artificial) agents which can be observed in a dialogue are not performed randomly but are part of the agent's *plan* (cf. Figure 52) to achieve some *goal* (Cohen 1997: 207; Huang et al. 2001: 888). Speakers generate contributions to the dialogue based on their plan (cf. Section 11.1.4)

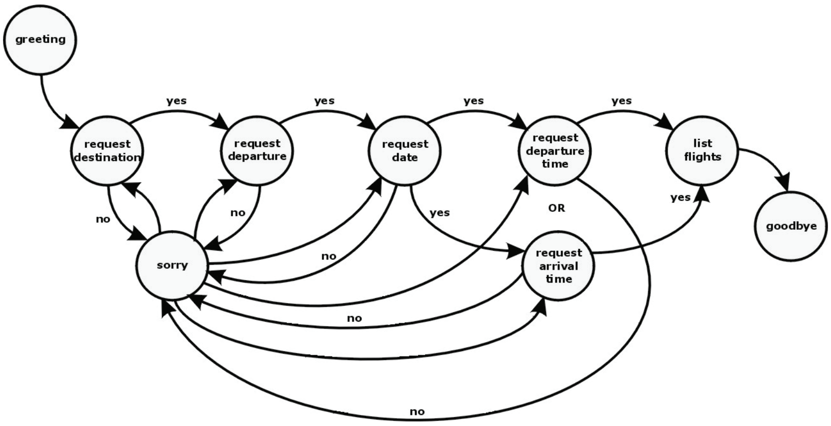

Figure 51. A finite-state model of a dialogue structure for flight inquiries (Androutsopoulos & Aretoulaki 2003: 640, Figure 35.3).

while listeners have to infer this plan from what they hear (and see), i.e. from the speaker's utterances and his or her non-verbal (communicative) actions (Cawsey 1998: 116f.; Jurafsky & Martin 2008: 850). The user's dialogue acts (e.g. a learner's request for the agent's assistance with drawing a syntactic tree diagram) are integrated into the evolving plan underlying the interaction, which may also involve physical actions in the domain (e.g. the agent draws a part of the tree and the user finishes it) (McTear 2002: 136). If a particular dialogue act is observed in the user's input, a plan or subplan underlying this act is constructed to identify possible goals of the user which he or she is trying to achieve by performing the act (cf. Chapter 12.2.1) (Allen 1994: 481–490; Carberry 2001: 32ff.; Huang et al. 2001: 890f.; Jurafsky & Martin 2008: 851ff.).

Plans show a hierarchical decomposition of actions into subactions and indicate the temporal ordering of the actions to be carried out to achieve goals and subgoals on each level of the hierarchy (Allen 1994: 481ff.; Tate 2001: 653; Russell & Norvig 2003: 422f.). The actions in a plan are conceived of as *plan operators* with associated pre-conditions and effects (Cawsey 1998: 88; Jurafsky & Martin 2008: 851f.). Some example operators corresponding to plan actions at different levels of detail in Figure 52 are provided below:

```
TAKE-TRIP(Actor,Vehicle,Dest)
Preconditions: DESTINATION(Vehicle,Dest)
Effects: IN(Actor,Dest)

BUY-TICKET(Actor,Ticket-office,Ticket-for(Vehicle))
Preconditions: HAS(Actor,Price(Ticket-for(Vehicle)))
Effects: HAS(Actor,Ticket-for(Vehicle))

PAY(Actor,Clerk,Price(Ticket-for(Vehicle)))
Preconditions: HAS(Actor,Price(Ticket-for(Vehicle)))
Effects: HAS(Clerk,Price(Ticket-for(Vehicle)))
```

Plan-based dialogue modeling is useful for collaborative problem solving (Ferguson & Allen 1998; Allen et al. 2001b; Ferguson & Allen 2007), which, among other things, involves the constant exchange of information between the participating agents (both artificial and human) in order to report on the current state of the joint problem solving and to supply information

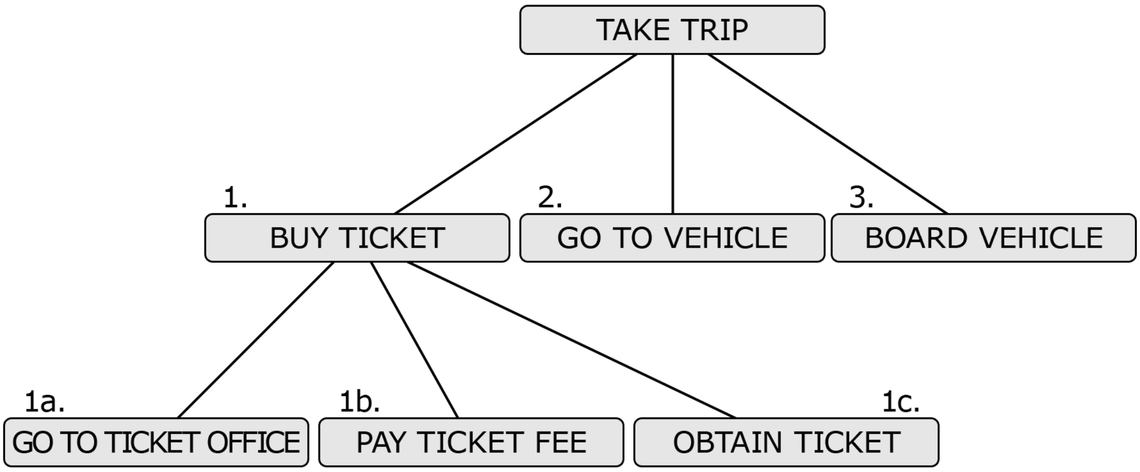

Figure 52. A plan showing the hierarchical decomposition of the high-level action/goal TAKE TRIP into successively more specific subactions/subgoals. Adapted from Allen (1994: 481, Figure 15.7). The numbers in the diagram indicate the temporal ordering of the actions on each level.

required by the other agent.[156] However, if a dialogue agent fails to identify the communicative act performed by the user, it may misconstrue his or her plan (McTear 2002: 136f.). With respect to system performance, it should be noted that plan generation and recognition are both computationally expensive processes (Cohen 1997: 209; Jurafsky & Martin 2000: 738; McTear 2002: 137).

While both the complexity of the models and the flexibility of the interaction increase from the finite-state through frame-based to plan-based approaches to dialogue modeling (Allen et al. 2001a), none of them covers the full range of human dialogue behavior. Therefore, human-machine dialogues in (spoken) natural language remain fairly restricted, with respect to both topics and language.

11.3.3.2.3 User Model

> *The idea that computers should adapt to the characteristics of their human users has become a major concern in human-computer interaction (HCI) as well as AI [(artificial intelligence)] research (...). It has been suggested that computers might be able to behave more co-operatively – and thus more usefully – if they had knowledge of the characteristics of their users and could use this knowledge to adapt their dialogues accordingly. (McTear 1993: 158).*

Dialogue agents require knowledge about the user to support their internal processing and to be able to participate effectively in agent-user conversations. This knowledge is collected in a process called *user modeling*,[157] represented and accumulated in a *user model*,[158] and applied to adapt the agent's communicative and other behaviors to a particular user, often dynamically while he or she is interacting with the agent (Wahlster 1988; McTear 1993; Zukerman & Litman 2001; Jameson 2003; Johnson & Taatgen 2005).

---

156 An elaborate list of activities of agents involved in collaborative problem solving was provided by Ferguson and Allen (2007).

157 User modeling will be discussed in detail in Chapter 12.2.1.

158 The equivalent component of a pedagogical agent is the student model (cf. Chapter 12.2).

The information in a user model may both remain fairly stable across interactions (e.g. age, gender, personality, skills, and vocabulary) and change as the dialogue progresses (e.g. goals, beliefs, preferences, emotions, and task parameters) (Cole et al. 1995; McTear 2004: 124).[159] User models may range in sophistication from a broad conception of a prototypical or stereotypical user of the agent through separate models for different classes of users (e.g. users with different expertise levels or from different age groups or sociocultural backgrounds) to personalized models of individual users which are dynamically built and incrementally refined over a history of conversations between the user and the agent (McTear 1993: 160ff.).

Dialogue agents can use the information in the user model to adapt their behavior and improve their linguistic processing when dealing with a particular user. For example, speech recognition and language understanding can exploit knowledge about how users with different backgrounds and (levels of) expertise express themselves (word choice and constructions) and what meanings they typically (intend to) convey (Cole et al. 1995). In dialogue management, a different, more restrictive dialogue strategy tending toward system initiative may be adopted if information collected about the performance of the interpreter on the spoken input of a particular user consistently indicates poor performance within and across dialogues (Zukerman & Litman 2001: 148; Jameson 2003: 308). The response generator (cf. Section 11.3.3.3) may use information in the user model to generate and deliver responses that are suitable (with respect to content, terminology, and complexity of constructions) for users with a certain background, domain knowledge, goals, preferences, misconceptions (cf. Chapter 12.2.3.2), and technological literacy (McKeown & Moore 1997: 184). For delivery of the message, the user model could guide the process of selecting and modulating a (synthetic) voice that the user finds familiar or pleasant and that provides an articulation of the message which is appropriate for the content of the message and the communicative context.

11.3.3.2.4 Error Handling

Given the many possibilities for non-understanding and misunderstanding in human-computer dialogues, the problem of failure must be dealt with on a regular basis at all levels of dialogue processing, from recognizing the spoken words up to interpreting the speaker's communicative intentions behind his or her utterance (McTear et al. 2005: 249f.). As a result, intelligent *error handling strategies* become a key factor for successful dialogue-based interaction (Huang et al. 2001: 937–941; McTear 2004: 353–362; McTear et al. 2005).[160] Effective error handling involves three stages. *Error prevention* is the task of anticipating potential errors in interactions with users while the dialogue agent is being developed, and designing the agent to prevent these errors in the first place, for example by phrasing its prompts in a way that limits the range of possible user responses or by guiding the dialogue along paths that are within the agent's capabilities (McTear et al. 2005: 250). *Error detection* is the process of constantly monitoring the exchange over time to pick up any signs of problems in the dialogue. Errors may be detected immediately when they occur (*early detection*) or only later during the dialogue (*late detection*) (p. 250). The former happens with non-understandings, whereas the latter takes place

[159] For examples of information that may be included in user models of agents, see Chapter 12.2.1.

[160] McTear et al. argued that "errors are a *natural occurrence* in spoken communication, both in human-machine dialogues as well as [sic!] in human-human communication. (…) [W]hat is required is an approach that treats errors, at whatever level, as *unavoidable* and that has methods for detecting and dealing with miscommunication when it occurs" (McTear et al. 2005: 249f., emphasis added).

when an utterance has been misunderstood (Skantze 2003). The results of monitoring the dialogue can also be used for *error prediction*, i.e. for anticipating and preventing potential future problems in the ongoing exchange based on observations made up to the present point (McTear 2004: 360f.). Finally, *error recovery* involves deciding on an appropriate course of action to respond to and recover from errors that have been detected in a dialogue. In general, there may be several options (McTear et al. 2005: 252):

- Ignore the error and hope for the best.
- Ask the user to repeat or rephrase his or her query.
- Perform explicit or implicit verification of what the agent thinks it understood.
- Try to correct the error alone using internal knowledge and knowledge of the context.
- Switch to another input channel (cf. Section 11.2).[161]

The best way for an agent to handle a given non-understanding or misunderstanding depends on the nature and severity of the error, the situation in which it occurred, the user involved, and other factors. In general, dialogue agents should be careful who they blame for problems in the conversation. Nass and Brave (2005: 173f.) reported on an experiment in which simulated misrecognition errors occurred during the interaction of the participants with a (fake) voice interface. The system either blamed itself or the user for the errors. In both conditions, the target of the blame (the system or the user) was apparent from the wording of the error messages. The results showed that blame for speech recognition errors was not well received by participants: as suggested by their ratings, they perceived the system which blamed them as less likeable and the interaction as more frustrating compared to the self-blaming interface. Hence, use of modest (self-blaming) language contributes to positive user attitudes toward an interface. But like people with a strong tendency to blame themselves for errors, a self-blaming interface is also seen as less competent. When an interface criticizes its own performance, users take this self-blame at "face value," without looking deeper (Nass & Brave 2005: 174).

If a conversational agent that blames the user is perceived as less likeable and a self-blaming agent as less competent, are there other alternatives? One possibility is to use neutral messages that do not blame anyone. However, people commonly have a *self-serving bias*, i.e. a tendency to claim success for themselves and hold others responsible for failure. Therefore, they would perceive the agent as blaming itself or, worse, interpret the agent's behavior as an attempt to sneak out of responsibility (Nass & Brave 2005: 175). Another possibility mentioned by Nass and Brave is for the agent to put the blame on an acceptable *scapegoat*, such as the high level of background noise or the low-quality microphone in the event of speech recognition errors. But scapegoating can backfire on the agent as well because people tend to think that any individual apart from themselves is to blame for his or her own failures, rather than someone or something in the environment, and therefore may regard scapegoating as an obvious attempt to escape from responsibility (p. 175). However, when users perceive that they form a team with the agent and depend on their agent partner (as the agent does on them) for the success of the team, they may extend the self-serving bias to include the agent. As a result, they may blame neither themselves nor the agent for failure but may be more willing to accept a scapegoat outside the

[161] For example, the agent could ask the user to type rather than speak difficult-to-pronounce words, including proper names, foreign-language terms, word creations, scientific and technical terms, or the user could use pointing gestures to disambiguate a spoken command, as in the famous "Put that there" system (Bolt 1980).

team (p. 175). This idea was supported by an experimental study conducted with a voice-based car interface (Jonsson et al. 2004).

While sophisticated error handling is important, preventing interaction problems is generally preferable to having to recover from them (Nass & Brave 2005: 177). One way to achieve this takes advantage of the human tendency toward *entrainment* or *mirroring*, i.e. toward adopting elements and patterns of the behavior of others ([INT 100]). Over time, people will gradually take up both the words and the grammatical structures which they hear from their interlocutors in dialogue, without thinking or even being aware of it (Nass & Brave 2005: 177f.). These principles of *lexical*[162] and *syntactic alignment* have been shown to apply not only when the interlocutor is human but also when it is a machine (Branigan et al. 2003; Pearson et al. 2006). Conversational agents can take advantage of these alignment tendencies to guide the user toward using words and constructions which the agent can handle (Nass & Brave 2005: 177f.).

Pearson et al. (2006) demonstrated that the language of users in human-computer dialogues is influenced by what they believe about and expect from the machine. Thus, if a conversational agent insinuates limited linguistic ability (for example by using a non-human embodiment (cf. Chapter 9.1), the user may be inclined to adapt his or her language behaviors to match the language of the agent (Nass & Brave 2005: 178). Pearson et al. found that this inclination is indeed stronger when users deem a system unsophisticated and less capable than when they think that it is sophisticated and capable (Pearson et al. 2006). Low sophistication or capability is suggested by cues including use of synthetic speech, slowness of speech, low-fidelity speech, ungrammatical speech, lack of multisyllabic words, simple sentence constructions, and self-blaming for failure (see above). When users perceive these cues, they are likely to simplify their language and to articulate their speech more clearly, much like they do when talking to foreign speakers or children (Nass & Brave 2005: 178). However, portraying an agent as 'stupid' is not always appropriate. In particular, this strategy will not be effective when the agent's role relies on users' perceptions of the agent's competence. For example, pedagogical agents as coaches (cf. Chapter 16.2.3) or presenters of information (cf. Chapter 12.3) will not be taken seriously if learners think that these agents are not knowledgeable about their subjects (cf. Chapter 7.2).

The willingness of users to adapt to the limitations of a conversational agent can also be influenced by social identification. Given that humans generally tend to socially identify with both people and technologies which they perceive as similar to themselves, agents can be designed with matching gender, personality, accent, and so on in order to induce users to make more of an effort to adapt to the agent's limitations (Nass & Brave 2005: 178). Another strong incentive for users is *reciprocity* (pp. 178f.): the perceived effort of an agent to understand the user makes the latter willing to reciprocate by making it easier for the agent to understand them (Fogg 1997; Moon 1998; Katagiri et al. 2001). The perception of effort can be supported by signaling to the user that he or she has the agent's full attention. Embodied conversational agents (cf. Section 11.3.4) can use eye gaze to indicate that they are paying attention to the user (cf. Chapter 10.1.4). Furthermore, they can mirror the user's choice of words and phrases in their responses. Finally, conversational agents can perceptibly pause to 'think' about a complex or important user request before they answer, thus showing the user that they deem the request worthy of their attention and effort (Nass et al. 2004; Nass & Brave 2005: 179).

[162] Nass and Brave used the term "semantic alignment" (Nass & Brave 2005: 177).

11.3.3.3 Response Generator

Apart from the performance of the interpreter and the dialogue manager, the effectiveness of a dialogue agent depends critically on its capability to respond to the user with a meaningful and intelligible spoken or textual message. The component of a dialogue agent that generates and delivers the agent's contributions to the dialogue is called the *response generator* (McKeown & Moore 1997; Huang et al. 2001: 894–901; McTear 2002: 124–127; McTear 2004: 99–104; Jurafsky & Martin 2008: 825ff.).

The response generator uses a natural language generation (NLG) component (cf. Section 11.1.4),[163] which generates linguistic outputs from a representation built by the dialogue manager (Huang et al. 2001: 895–899; McTear 2004: 99; Jurafsky & Martin 2008: 825f.). If the agent delivers its messages using speech, the generation stage must be coupled with a text-to-speech or concept-to-speech synthesizer (cf. Section 11.1.2), which produces the spoken output either from a text or from an enriched linguistic description of the message (Huang et al. 2001: 899; McTear 2004: 102). Response generation involves carefully selecting, organizing, and expressing content in a way that is tuned to the context (including the current dialogue context as well as the larger sociocultural context) (McTear 2002: 124) and conveys the agent's state, its personality, and its status and relationship with respect to the user.

Virtually all NLG systems described in the literature generate longer passages of text rather than utterances for a spoken dialogue (Theune 2003: 3). However, spoken language generation in a dialogue goes beyond the production of written monologues in several ways. First, the agent has to determine what to say, how to say it, and how to pronounce it (i.e. generate the prosodic structure of the utterance) (McKeown & Moore 1997: 183). As discussed in Section 11.1.2.3, this calls for a closer integration of language generation and speech synthesis in a concept-to-speech approach (Huang et al. 2001: 899). Furthermore, the selection, organization, and expression of content must be tuned to the larger context of the dialogue, which includes the history of previous interactions, the current dialogue state, and information about the user (McKeown & Moore: 1997: 183f.; McTear 2002: 124). In particular, the following guidelines need to be observed (Cole et al. 1995; Theune 2003: 4):

- Communicate only information that pertains to the user's goals and knowledge.
- Relate new to previously communicated information.
- Select appropriate expressions for referring to the entities under discussion.
- Avoid unnecessary repetitions.
- Use vocabulary and constructions that are comprehensible and fit the spoken context.
- Tune the wording of messages to those of the human dialogue partner (entrainment).
- Consider the user's cognitive and emotional state as well as his or her individual and sociocultural background.

In general, the appropriate form of language for spoken dialogue settings, including aspects like utterance length, complexity of constructions, and choice of words, is a problem that requires further research. However, the properties of spoken utterances generated for a dialogue will certainly differ from those of sentences produced for a text. For example, spoken messages will have to be shorter and less complex to facilitate comprehension (McKeown & Moore 1997: 187). Furthermore, they should communicate the agent's emotional, mental and physical state, its personality (through message structure, wording, voice identity, etc.), and its status (lower,

[163] See Theune (2003) for a review of NLG systems for response generation in dialogue.

equal, or higher) and relationship with respect to the user (for example, strangers and close friends should be addressed in different ways). Finally, for embodied conversational agents (cf. Section 11.3.4), the spoken-response generator has to be extended to a multimodal response generator, which additionally produces instructions for the facial and body movements that should accompany the agent's verbal message.

Until recently, research and development have largely focused on the input side of dialogue agents (McTear 2004: 102). Many practical dialogue systems have adopted a simple approach to response generation, drawing their outputs from a database of canned text strings that may contain empty slots to be filled in with information to be presented (Theune 2003: iii). While canned responses or, slightly more advanced, filled templates (cf. Section 11.1.4.3) are suitable for applications in which a dialogue agent can respond with short and simple messages whose wording is relatively fixed, others, such as tutoring (cf. Chapter 12.3.9), require more flexible language generation that is sensitive to the context (Theune 2003: iii + 4; McTear 2004: 102).

11.3.3.4 State of the Art

Dialogue agents have reached a state of development that allows their deployment for limited practical tasks in small and well-defined domains. Doyle (1999) argued that the nature of the interaction between user and computer determines the success of dialogue agents. He identified the following scenarios in which dialogue agents can be successful (Doyle 1999):

- *Uncertain vs. decisive*. Dialogue agents can assist users who are uncertain about particular choices with suggestions, advice, and information.
- *Satisfying vs. correct*. Dialogue agents can increase user satisfaction by making tasks more pleasant (e.g. providing verbal scaffolding in a foreign language learning task) or by convincing the user of the quality of an outcome (e.g. praising a learner who has successfully completed a difficult task).
- *Dialogue vs. command-driven*. Dialogue agents can engage users in conversations that help them to clarify their thoughts, gain insights, acquire skills, etc., facilitated by explanations, clarifications, suggestions, and instructions provided by the agent (cf. Chapter 12.3.9).
- *Social vs. natural*. Dialogue agents can build relationships with users by enriching task-related dialogues with exchanges of a social nature, such as small talk (cf. Chapter 10.3).

However, dialogue agents still have considerable failure rates, not only in recognition and understanding but also when it comes to the successful completion of interactions. Some challenges for research on dialogue agents include (McTear 2002: 161f.):

- More robust speech recognition (cf. Section 11.1.1.3);
- More natural and expressive synthetic speech (cf. Section 11.1.2.4);
- Closer integration and symbiosis of processing components;
- More sophisticated approaches to dialogue management;
- Use of statistical methods and machine learning techniques (cf. Chapter 12.1.1.2);
- Support for conversations in multiple languages;
- Enabling spoken dialogue in web-based applications;
- Move from disembodied to embodied conversation.

11.3.4 Embodied Conversational Agents

> *Interaction between people is the sum of the activities of the involved subjects, like speaking, gesturing, listening, watching. If a system simulates physical objects or characters, and is able to perceive human actions and intentions, people can interact with them [(i.e. the physical objects and characters)] using the ordinary schemes that they use in real life, and this is the key to a new level of satisfaction. (Valli 2004: 9).*

Human-human face-to-face conversations are 'embodied,' combining appropriately modulated speech with gaze, gestures, facial expressions, body posture and movement (cf. Chapter 10.1). Embodied conversational agents (Cassell et al. 2000a; Cassell 2001) with animated humanoid bodies (cf. Chapter 9.1) and equipped at least with auditory and visual input channels, i.e. the ability to process inputs combining speech, gestures, facial expressions, gaze patterns, etc., are perceptual multimodal interfaces (cf. Section 11.2) with the ability to engage users in multimodal dialogues involving synchronized verbal and non-verbal behaviors that realize different conversational functions (Bente et al. 2000; Bickmore 2004):

- *Modeling functions* serve to demonstrate a sequence of motor actions when the dialogue is concerned with particular motor skills.
- *Propositional functions* represent information related to the content of the conversation by means of meaningful speech as well as non-verbal behaviors that complement, supplement, or substitute for aspects of the speech content.
- *Interactional (or 'envelope') functions* include turn-taking, initiation and termination, engagement (to check if the other interlocutors are still "on board"), and framing functions (to signal shifts from e.g. task-related talk to small talk), among others. Their purpose is to regulate the flow of the dialogue.
- *Attitudinal functions* indicate what the communicator[164] thinks of the other participants in the conversation. A positive attitude can be conveyed by non-verbal immediacy behaviors (cf. Chapter 10.3.1).
- *Affective display functions* convey the communicator's affective state, including mood and emotion (cf. Chapter 13.1). Affective displays can be spontaneous expressions of internal state, or they may be used deliberately to achieve certain ends (cf. Chapter 13.2).
- *Relational functions* serve two purposes: they either signal how the communicator currently views his or her relationship with the other participants, or they are involved in giving this relationship a particular (desired) direction. Relational behaviors can make users feel good and more motivated, which has positive effects on task performance (cf. Chapter 7.16).

The components of an embodied conversational agent are shown in Figure 53 (Cassell et al. 2000b: 40–46; Cassell 2001: 74f.; Cassell 2007: 359f.).[165] The agent accepts user inputs as combinations of speech, facial expressions, gaze, gestures, postures, and locomotion via its *input devices*. Cameras are used to watch the user's face, gestures, postures, and movement while microphones capture his or her speech (Cassell 2007: 359). This stage involves a speech

---

[164] The term 'communicator' is used here and in the following in a more comprehensive sense than 'speaker' to refer to an individual that makes use of speech as well as non-verbal communicative behaviors to communicate messages.

[165] The architecture of an embodied *interactive* software agent, which was presented in Figure 16, is both more general and more comprehensive than this architecture and the architecture of a dialogue agent (cf. Figure 49), although it includes elements of both.

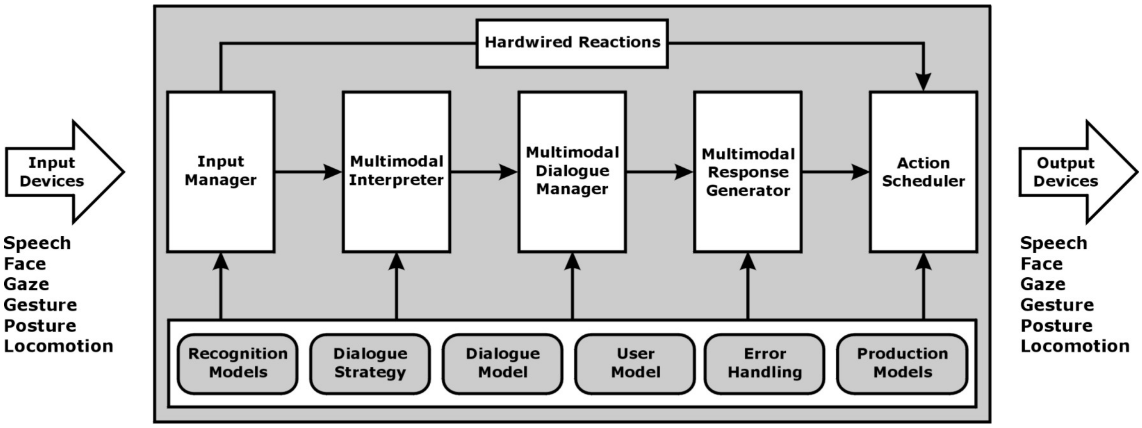

Figure 53. The architecture of an embodied conversational agent. Based on the architecture presented by Cassell et al. (1999, Figure 1; 2000b: 43, Figure 2.2) and Cassell (2000b: 75, Figure 2; 2001: 75, Figure 3; 2007: 359, Figure 17.6).

recognizer as well as technologies for the recognition and tracking of faces, facial expressions, lip, eye and head movements, gestures, postures, and locomotion. The data collected by all available input devices is sent to the *input manager*, which transforms it into a representation that can be used in further processing, and determines the necessity of an immediate response vs. additional multimodal dialogue processing (Cassell et al. 2000b: 42f.). Certain unimodal inputs, for example when the user is moving in front of the camera or entering or leaving the agent' field of view, trigger instant responses (e.g. tracking the user's locomotion with the eyes) from the agent that are handled by a *hardwired reactions* module (p. 43). The task of the *multimodal interpreter* is to integrate the information obtained from the different input channels into an overall interpretation of the user's verbal and non-verbal actions, which may consist of both *interactional* and *propositional* communicative acts (p. 44, cf. Section 11.2).[166] At the heart of the architecture is the *multimodal dialogue manager*, which processes interactional and propositional communicative acts in relation to the situation and content of the conversation, respectively, handling errors, updating the dialogue model and the user model, and deciding on an appropriate response in line with the agent's dialogue strategy (p. 44). This response is designed by the *multimodal response generator* as a sequence of actions (speech, gesture, etc.) to be performed by the embodied agent in order to achieve goals related to the communication or the task (pp. 44f.). Low-level output actions are scheduled by the agent's motor controller, which is called the *action scheduler* (p. 45). Finally, the *output devices* of the architecture render the agent's verbal and non-verbal behaviors as combinations of animation and recorded or synthetic speech. The processing modules of the architecture draw on a set of resources (shown at the bottom of Figure 53), including recognition and production models for speech, faces and facial displays, gaze patterns, and different kinds of gestures, body postures, and movement; a dialogue strategy which determines the allocation of control in the conversation; a continuously updated model of the ongoing dialogue; a profile of the user; and strategies for dealing with errors in the interaction.

Although the architecture in Figure 53 shows a sequence of processing steps from perception to production, Cole et al. argued that embodied conversational agents actually have to be

[166] Propositional communicative acts add content to the conversation while interactional communicative acts are responsible for regulating the conversational process (Cassell et al. 2000b: 44).

capable of real-time, *simultaneous* interpretation and production of auditory and visual conversational signals (Cole et al. 2003: 1401). While making its contribution to the dialogue, the agent must monitor and interpret the user's behavior through both the auditory and visual channels to pick up indicators of the user's reaction (e.g. agreement, confusion, or wish to interrupt), and while the user has the floor, the agent must interpret the combined verbal and non-verbal input from the user and at the same time respond with appropriate audio-visual feedback through its animated persona (p. 1402).

Embodied conversational agents were originally used as a scientific instrument to study phenomena of face-to-face conversation, but they are also increasingly designed and deployed as tools for everyday people to interface with computers (Cassell 2007: 368), facilitating access to information and services, serving as advisors or representatives, or playing various roles in interactive learning environments (Johnson et al. 2000; Ryokai et al. 2003; Massaro 2005; Wise et al. 2007), among other things. One of the most faithful simulations of human face-to-face conversational behavior (given the current state of technology) has been implemented in REA (for "Real Estate Agent"), a female humanoid embodied conversational agent in the role of a realtor that engages its human interlocutors in a face-to-face house-selling dialogue (Cassell 2000b; Cassell et al. 2000b; Cassell 2001; Cassell 2007).

11.4 Summary

This chapter discussed the design of the conversational capabilities of embodied interactive software agents. The major theme was the (embodied) conversational agent, which has the ability to participate verbally (and non-verbally) in dialogues with the user that rely on language or integrate it with modes of non-verbal communication, which were discussed in Chapter 10.1. Conversational agents have to be able to recognize and interpret, generate and express spoken or written messages in one or multiple languages. Hence, Section 11.1 reviewed the state of the art in different human language technologies, including speech recognition (cf. Section 11.1.1), speech synthesis (cf. Section 11.1.2), natural language understanding (cf. Section 11.1.3), and natural language generation (cf. Section 11.1.4). In addition, agent-user conversations may involve several channels (modalities) through which the participants may exchange messages. Multimodal interfaces, discussed in Section 11.2, provide the technologies for processing combined user inputs from at least two input modes in a coordinated fashion and for producing multimedia responses as output.

Equipped with the necessary background on the technologies for processing natural language and multimodal inputs, several types of increasingly sophisticated conversational agents were discussed in Section 11.3. First, a summary of several fundamental principles (turn-taking, deixis, cooperativeness, and politeness) that allow human-human (and, by analogy, human-agent) conversations to be conducted smoothly and effectively was provided in Section 11.3.1. Next, chatbots were introduced, i.e. computer programs that simulate text-based conversation, or chat, applying pattern-matching techniques to associate a given user input with a pre-scripted response (cf. Section 11.3.2). Dialogue agents (cf. Section 11.3.3) use more sophisticated speech and language processing technologies to enable more flexible and purposeful agent-user dialogues. These agents consist of an interpreter, a dialogue manager, and a response generator, which are supported by a set of resources for dialogue processing, including the dialogue strategy, the dialogue model, the user model, and mechanisms for error handling. Issues, options, and techniques involved in designing these processing stages and resources were reviewed in Section 11.3.3. Finally, Section 11.3.4 discussed embodied conversational agents, which integrate multimodal input processing technologies and an articulate face and body to enable conversations between agents and users in which both parties can use combined verbal and non-verbal communicative behaviors that serve various functions in conversations. The architecture of an embodied conversational agent was presented and its elements were discussed.

12 Expertise

People will be motivated to interact with agents which they perceive as being able to give them something useful. Often, these perceptible qualities of an agent (cf. Chapter 7.2) include some form of task-oriented or social expertise that may help the user in his or her current situation (Hayes-Roth 2004: 454). *Expertise* can be defined as a quality of individuals or systems that possess substantial knowledge or ability in a certain area, or *domain*, which enable them to make pertinent contributions to a situation, including advice, information, judgment, and action (Karlsen & Karlsen 2007, [INT 101]). Hence, agents communicating that they have some kind of expertise must be able to justify users' perception of that expertise by delivering results and performance which the users expect. The expertise of agents may be of various kinds:

- *Encyclopedic*. The agent can provide relevant information in sufficient breadth and depth when needed or requested.
- *Instructional*. The agent can design and implement instruction for both human and agent learners that is challenging, understandable, comprehensive, and in line with principles and best practices of instructional design (cf. Chapter 2.2).
- *Problem-solving*. The agent has the ability to find solutions to problems which are presented to it or which it identifies or creates on its own.
- *Social*. The agent has competencies in (embodied) communication and relationship building.

Pedagogical agents add value (cf. Chapter 7.1) to learners' experience if they not only possess these kinds of expertise but also advertize them in such a way that learners perceive and make use of them. Depending on their instructional role (cf. Chapter 16), pedagogical agents rely on their encyclopedic, instructional, problem-solving, and/or social expertise to:

- Provide information about the subject matter;
- Create problems for the learner to solve;
- Solve problems on their own and walk the learner through their solutions;
- Analyze the learner's behaviors and mistakes;
- Determine the learner's understanding of a topic;
- Adapt subject matters and instructional methods to the learner;
- Communicate effectively with the learner;
- Build productive and lasting relationships with the learner.

This list is a superset of the arsenal of instructional activities that is available to traditional (intelligent) computer-based tutors (cf. Chapter 2.3). Each of these activities poses considerable challenges for human instructors, and even more so for pedagogical agents, since they require capabilities including logical deduction and inference, creativity, learning, decision making, social and emotional intelligence (cf. Chapter 13.4), and language processing (cf. Chapter 11), which have been developed for intelligent machines in the field of artificial intelligence since the 1950s.

The expertise of pedagogical agents is represented by a set of both pre-built and dynamically constructed knowledge sources, including a *domain model* representing an agent's knowledge of its subject domain, a *student model* capturing the agent's evolving view of the learner and his or her development, and an *instructor model* embodying the agent's instructional expertise.[167] Each of the highlighted models is covered in a separate section of this chapter, starting below with the domain model.

12.1 The Domain Model

One of the fundamental insights of artificial intelligence research is that intelligent behavior of any kind presupposes substantial knowledge (Cawsey 1998: 3). While the term 'knowledge' is commonplace and seems intuitively clear, a conclusive definition has eluded philosophers since the ancient Greeks (Brachman & Levesque 2004: 2), when Plato first postulated that for statements to become knowledge they have to be 'justified,' 'true,' and 'believed' ([INT 102]).[168] It is neither possible nor intended here to provide the final definition. For the purpose of the present work, *knowledge* can be defined as "information combined with experience, context, interpretation, and reflection (…) that is ready to apply to decisions and actions" (Davenport et al. 1998: 43). Knowledge can be classified along several dimensions, which are described in Table 17.

It is important to distinguish knowledge from another related notion, *belief*, which concerns the individual's mental perspective on reality, in particular the propositions that he or she holds to be true ([INT 103]; [INT 104]). Given Plato's definition of knowledge as "justified true belief," at least two conditions have to be fulfilled for a belief to become knowledge: it must be consistent with reality (i.e. true), and there must be a plausible reason (or justification) for the belief (e.g. empirical evidence or a sound argument) (Russell & Norvig 2003: 343; [INT 103]; [INT 104]).

Knowledge is the key ingredient when developing the cognitive component[169] of pedagogical agents and other intelligent systems. The cognitive component or 'brain' consists of the set of computational models and procedures which through their interplay enable behavior that manifests the expertise of the agent or system. All agents require some sort of representation that accumulates, structures, and encodes the knowledge which the agent has about its area of expertise (or *domain*). The agent's *domain model* contains (expert) knowledge about:

- *Objects* (physical entities and abstract concepts);
- *Facts* (relationships between or among concepts or entities);

---

[167] While these components are similar to those found in an intelligent tutoring system (cf. Chapter 2.3.3), this is not intended to suggest that 'tutor' is the only role that pedagogical agents can play in interactions with learners (although tutor agents have been built in many research projects). In fact, there are various other roles for pedagogical agents, and some examples will be discussed in Chapter 16. To avoid confusion, the term "instructor model" will be used instead of "tutor model" in this chapter.

[168] For a long time, Plato's view of knowledge as "justified true belief" enjoyed considerable popularity. However, in the 1960s, it was challenged by the American philosopher Edmund Gettier, who gave several counterexamples illustrating beliefs that, while true and justified, intuitively would not be regarded as knowledge (Gettier 1963, [INT 103]).

[169] Cognition is not the only mental information processing system that is vital to the functioning of an intelligent agent. As argued in Chapter 6.4, cognition is both influenced by and influences another system, affect, which provides rapid and continuous positive or negative assessments of events and situations. Emotion and personality are two central elements of the affective system of an agent. They will be discussed in the next chapter.

Table 17. Dimensions of knowledge.

Declarative ←→	**Procedural**
Declarative (or Propositional) Knowledge	**Procedural Knowledge (or Know-How)**
Knowledge about the truth (or falsehood) of a given statement (or proposition) ([INT 105])	Knowledge about how to perform an activity or accomplish a task ([INT 106])
A Priori ←→	**A Posteriori**
A Priori Knowledge	**A Posteriori (or Empirical) Knowledge**
Knowledge acquired independently of experience in the world ([INT 103])	Knowledge acquired from experience in the world (e.g. through experimentation, observation, etc.) ([INT 103])
Factual ←→	**Inferential**
Factual Knowledge (or Know-What)	**Inferential Knowledge**
Knowledge consisting of verified statements about things that are the case or have occurred (i.e. facts)	Knowledge derived by reasoning about existing facts or other inferred knowledge ([INT 105])
Domain(-Specific) ←→	**Domain-Independent**
Domain(-Specific) Knowledge	**Domain-Independent Knowledge**
Knowledge particular to a pre-determined subject area (or domain) ([INT 107])	Knowledge that can be applied across domains ([INT 107])
Expert ←→	**Common Sense**
Expert Knowledge	**Common Sense Knowledge**
Knowledge extracted from individuals with special knowledge and skills in a particular domain ("domain experts") ([INT 101])	Knowledge that (almost) every human being is supposed to have ([INT 108])
Situated ←→	**General**
Situated Knowledge	**General Knowledge**
Knowledge that pertains to a specific situation; often part of language, culture, or traditions ([INT 102])	Knowledge that can be used in various situations
Explicit ←→	**Tacit**
Explicit (or Codified) Knowledge	**Tacit Knowledge**
Knowledge that can be articulated, formalized, stored, and transmitted to others ([INT 109])	Knowledge that cannot be articulated or recorded but has to be acquired through training or personal experience ([INT 110])

- *Events* (things that happen in a certain place at a certain time and for a certain period);
- *Procedures* (knowledge about how to do different things);
- *Knowledge* (*metaknowledge* about the amount, reliability, sources, etc. of knowledge).

All of these aspects pertain to the agent itself, its environment, or its task (role). To be able to navigate in a learning environment and manipulate or refer to its objects, a pedagogical agent has to know about the objects in the environment, their attributes and spatial properties. Task knowledge includes both know-how for performing tasks in the domain and knowledge that allows the agent to explain, give hints, and answer learners' questions about the steps involved in the relevant procedures (Rickel & Johnson 1999). Thus, the knowledge of a pedagogical agent has to be deeper than the knowledge of agents that only have to carry out domain tasks because it additionally has to be able to make the principles and relationships underlying its actions explicit to the learner (Johnson 1998; Johnson et al. 2000).

The construction of domain models for agents requires careful consideration and application of effective techniques for several knowledge-related activities, which are discussed in the subsections below:

- *Acquisition*. How to extract knowledge from individuals, data, and experience;
- *Representation*. How to organize, structure, and encode knowledge in order to facilitate acquisition and use;
- *Use*. How to apply the represented knowledge in problem solving.

12.1.1 Knowledge Acquisition

Building domain models is among the more challenging aspects of agent development. The desired knowledge is typically not available in a form which agents can use directly. In fact, it is normally buried in the minds of human domain experts, in textbooks, in databases or corpora, and has to be elicited, collected, analyzed, modeled, and validated ([INT 111]) so that it can be represented in a knowledge base[170] and used by artificial agents. This is the realm of *knowledge acquisition* (Marcus 2001; Potter 2001; Milton 2007), which has been defined as

> *the process of acquiring (either directly from a human or from some other source) that information, and its formalized structure, that will allow some particular task to be performed by a computer system. In AI [(artificial intelligence)], structured information of this sort is commonly termed 'knowledge' (...) (Potter 2001: 1).*

12.1.1.1 Knowledge Engineering

Traditionally, knowledge acquisition has been the task of *knowledge engineers*, professionals who extract domain knowledge from human experts and translate it into a knowledge base for an agent or other intelligent system (Cawsey 1998: 40–44; Studer et al. 1998; Russell & Norvig 2003: 260–266; Kendal & Creen 2007). Based on his or her inquiry into a particular domain, a knowledge engineer identifies important domain concepts and develops a formal model of the domain that captures the discovered concepts and their relationships (Russell & Norvig 2003: 260). Eventually, this model is implemented as the knowledge base of the agent or system.[171] A number of issues tend to make the knowledge engineering process difficult ([INT 111]):

---

[170] A *knowledge base* is a database in which domain knowledge can be collected and organized for later retrieval by humans and/or machines ([INT 112]). In the present work, the term is synonymous with "domain model" (cf. Chapter 11.1.3.2.4).

[171] Originally, knowledge engineers were specialists in the field of artificial intelligence concerned with the development of *expert systems* (Cawsey 1998: Chapter 3; Jackson 1999), i.e. software systems designed to aid in problem solving or decision making in a particular area of human expertise, such as medicine, business, and education, by making inferences from the knowledge of domain experts represented in the system's knowledge base as an extensive collection of IF-THEN rules (cf. Section 12.1.2.4) and (in some systems) frames (cf. Section 12.1.2.3) (Cawsey 1998: 41; Kleinedler 2001: 100f.; [INT 113]).

- Knowledge in the minds of experts is not directly accessible.
- Experts know a lot about their domain. How broad and deep should this knowledge be modeled for the target application?
- Much expert knowledge is *tacit* in nature, i.e. experts are not aware of everything they know. In general, tacit knowledge is very difficult to make explicit and to formalize.
- Expert time is valuable and expensive, but extracting knowledge from an expert can be a time-consuming process.
- The knowledge of individual experts about a given domain may be vast, but it is still incomplete. How does a knowledge engineer know what pieces are missing and where he or she can obtain them?
- Knowledge tends to move from current to out-dated to obsolete at an ever faster rate. Hence, regular updates of the knowledge base become essential.

Knowledge engineering processes typically involve the following steps (Russell & Norvig 2003: 261f.):

1. *Identify the task*. Different tasks require the representation of different knowledge in order for the agent or system to be able to produce solutions to specific problems with the help of the knowledge base.
2. *Assemble the relevant knowledge*. The knowledge engineer extracts knowledge about the domain from external sources (human experts, scientific publications, operating manuals, the World Wide Web, etc.) or from his or her own expertise.
3. *Decide on a representation vocabulary*. The knowledge engineer develops a vocabulary (*ontology*) of terms and constructs for describing the concepts, attributes, relationships, and constraints in the domain (cf. Section 12.1.2.5).
4. *Encode general knowledge of the domain*. The knowledge engineer codifies the basic rules of the domain, specifying the meaning of all the terms in the ontology. Domain experts can then check this specification to uncover misconceptions or gaps that need to be addressed by going back to the previous step.
5. *Encode a specific problem instance*. The knowledge engineer provides a description of a particular problem from the domain, about which the system will be queried.
6. *Pose queries to the inference procedure*. The knowledge base is tested by sending queries to the inference procedure, which works with the general domain knowledge and the particular facts of the problem to produce an answer.
7. *Debug the knowledge base*. Responses based on a given state of the knowledge base will be mechanically correct but may indicate gaps and errors in the represented knowledge that have to be corrected by the knowledge engineer.

A wide range of knowledge acquisition techniques have been developed to help knowledge engineers to elicit knowledge from an expert. The most common techniques are shown in Figure 54 and described below ([INT 111]):

- *Interviews* are one-to-one social interactions between an expert and a knowledge engineer in which the engineer asks the expert questions about his or her area of expertise (Potter 2001: 2f.; [INT 114], cf. Chapter 15.4.5.2). The interview can be *unstructured* (following some general themes but lacking a pre-defined structure), *semi-structured* (combining a pre-determined schedule with the possibility to explore topics outside the schedule that emerge during the interview), or *structured* (working through a pre-defined catalogue of questions).
- *Laddering* constructs, modifies, and validates *ladders*, i.e. hierarchical network diagrams (tree structures), such as taxonomies ([INT 115]; [INT 116]).

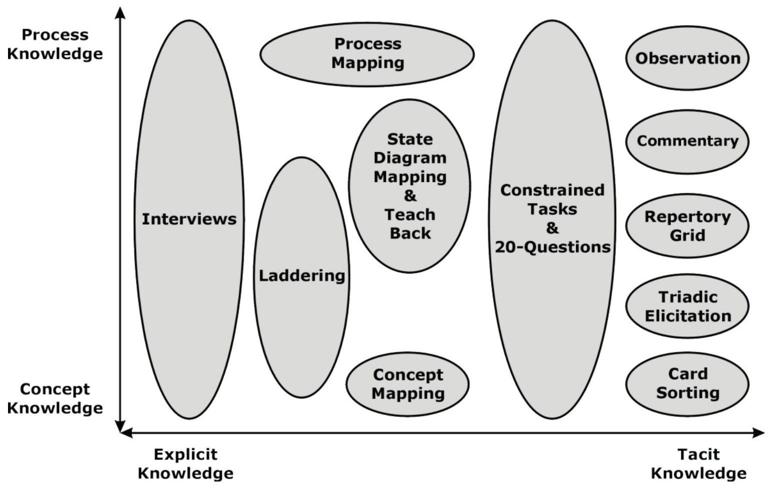

Figure 54. Knowledge acquisition techniques and the types of knowledge they are intended to elicit. Adapted from ([INT 111]).

- *Process mapping* constructs networks that capture the tasks involved in a process plus their associated inputs, outputs, resources, roles, and decisions ([INT 115]; [INT 117]).
- *State diagram mapping* builds, manipulates, and validates network diagrams that consist of nodes corresponding to different states of a domain concept, and links indicating events or processes that lead to a change of state ([INT 115]; [INT 117]).
- *Teach back* has the expert comment on descriptions of previously acquired knowledge given by the knowledge engineer in order to detect misunderstandings ([INT 114]).
- *Concept mapping* creates semantic-network-like structures (cf. Section 12.1.2.2) that capture domain concepts (nodes) and their relationships (labeled links) ([INT 115]; [INT 117]).
- *Constrained tasks*. The expert performs one of his or her daily tasks, but the time and/or information available for the task are restricted ([INT 118]).
- *20-questions*. The expert asks the knowledge engineer a series of yes-no questions to find out what domain-related concept the engineer is thinking about. The experts' questions and the order in which they appear reveal important concepts and properties of the domain in a prioritized order ([INT 118]).
- *Observation*. The knowledge engineer watches and documents the expert's performance of his or her daily tasks (cf. Chapter 15.4.5.3). It is also possible to make a video recording of the performance ([INT 114]).
- *Commentary*. The expert accompanies his or her own or someone else's performance of a domain task with a running commentary ([INT 114]).
- *Repertory grid.* This matrix-based technique involves representing a set of domain concepts and a set of attributes that differentiate between them in a two-dimensional grid; rating all concepts against all attributes on a scale; and grouping similar concepts and attributes based on cluster analysis. The resulting focus grid is explained to the domain expert in order to obtain feedback and further information about the groupings and correlations displayed in the grid (Crowther et al. 1995; [INT 119]).
- *Triadic elicitation* involves randomly selecting three domain concepts and asking the expert to pick two of them that he or she considers very similar and to articulate what makes them similar and what distinguishes them from the third concept. The goal of this technique is to elicit attributes that are otherwise difficult to obtain from the expert ([INT 120]).

- *Card sorting*. The expert is asked to repeatedly sort cards labeled with domain concepts into piles of cards that share certain properties. When asked to label each pile, the expert uses attributes and values that make up the properties of domain concepts ([INT 120]).

Despite support provided by automated tools (Marcus 2001: 429), knowledge engineering is still a time-consuming and error-prone, largely manual process, which may not be feasible for domains involving knowledge that is extensive, ill-defined, unknown, or dynamically changing. An alternative to crafting knowledge bases by hand is to enable agents and other systems to acquire or enhance their knowledge and capabilities on their own, with or without human supervision, by using machine learning algorithms.

12.1.1.2 Machine Learning

Machine learning (Langley 1996; Nilsson 1996; Mitchell 1997; Cawsey 1998: Chapter 7; Mooney 2003; Alpaydin 2004) is a field of study within artificial intelligence whose goal is to give artificial agents and other computer-based machines the ability to learn, i.e. to make changes to their internal structure, program, or data as they acquire knowledge and experience, which enable them to perform better in the same or similar situations in the future (Nilsson 1996: 1). 'Change' can either mean that existing systems or components are improved or that new ones with better capabilities are synthesized (p. 2). Paraphrasing Nilsson, machine learning techniques may be used to enhance the performance of all components of an artificial agent, including perception, environment modeling, planning and reasoning, and action computation (p. 2). Various reasons for exploring machine learning in agents and other intelligent systems can be identified (p. 3):

- The only way to define some tasks satisfactorily is by giving many examples pairing inputs and desired outputs. By learning the patterns implicit in these examples, agents can produce expected results for whole classes of inputs, including instances that were not part of the initial set of examples.
- Machine learning methods can uncover hidden relationships and correlations in data in processes called *data mining* (Hand et al. 2001; Hearst 2003) or *knowledge discovery in databases* (Frawley et al. 1992; Fayyad et al. 1996).
- It may not be possible to anticipate all aspects of the target environment while an agent is being designed. Machine learning can help agents to improve themselves "on the job."
- It may not be feasible to acquire the vast amounts of knowledge available or relevant in a particular domain by means of knowledge engineering. Equipping agents to learn this knowledge piece by piece on their own through automatic processing would be a more effective and efficient alternative to human encoding.
- Agents may be situated in environments that change with the progression of time. If they could adapt to such dynamic changes in their environment, there would be less need for developers to constantly modify their design.
- In realistic scenarios, both knowledge and the vocabularies (ontologies) for its description are growing and changing all the time. Constant hand-encoded updates to accommodate new or changed knowledge and vocabulary are not practical. The hope is that machine learning techniques can help agents to stay up-to-date.

Machine learning subsumes a wide range of computational techniques that have the common goal of improving the future performance of computer-based systems on particular tasks by acquiring knowledge from previous cases, experience, or exploration, among other sources

(Cawsey 1998: 181; Mooney 2003: 376). The acquired knowledge is represented in terms of various kinds of computational formalisms, including functions (relations mapping from inputs to desired outputs), logic programs and sets of rules, finite-state transition networks, grammars (cf. Chapter 11.1.3), and problem-solving systems (cf. Section 12.1.3) (Nilsson 1996: 5). Often, the term "machine learning" is reserved for techniques that use such declarative, symbolic formats to represent the results of learning, which excludes statistical or neural-network training algorithms that produce numerical output, like those used, for example, in speech recognition (cf. Chapter 11.1.1.2) (Mooney 2003: 376). However, the following list summarizes both symbolic and non-symbolic techniques for machine learning:

- *Rote learning ("learning by memorization")*. Learning that memorizes (i.e. organizes and stores) information obtained from the environment to retrieve it later, without processing this information further.
- *Induction*. Learning that generalizes from examples in order to be able to correctly process unseen future examples (Russell & Norvig 2003: Chapter 18). Inductive learning has been the focus of most efforts in the field of machine learning so far (Cawsey 1998: 144).

  Often, there will be multiple possible generalizations, or hypotheses, to choose from, that are all consistent with the example data. Therefore, inductive machine learning algorithms require a basis for making a choice among these alternative hypotheses. The term *bias* is commonly used to refer to a priori information that enables a learning agent to select one generalization over another if a choice cannot be made on other grounds (Dietterich & Kong 1995; Nilsson 1996: 10–13).

  All inductive learning algorithms depend on the size and the quality of the input data used for learning (the *training set*). To achieve good results in inductive learning, as demonstrated by the system's performance on a *test set* of unseen examples, the learning algorithm needs a large and representative collection of training data plus appropriate predictive features (i.e. features which are particularly suitable as indicators of desirable results). In other words, in practical learning situations, the particular machine learning technique chosen may be less important than the quality of the data available for training (Cawsey 1998: 164f.).
- *Deduction*. Learning that reasons about existing knowledge to derive 'new' facts that can be concluded from it based on principles of logic. For example, if both $A = B$ and $B = C$ are known, $A = C$ can be deduced (transitivity).
- *Supervised learning*. Learning an unknown function from a set of examples (supplied by a 'teacher') that pair inputs with desired outputs of the function (Russell & Norvig 2003: 650).
- *Unsupervised learning*. Learning an unknown function from a set of input patterns but without information about which outputs are desirable or correct. Agents learning entirely without supervision do not know which actions are 'correct' or which states are 'desirable' and thus are unable to learn which actions to take or which states to achieve (Russell & Norvig 2003: 650).
- *Reinforcement learning*. Learning in which the agent is not explicitly taught what to do but receives (positive or negative) feedback on its performance that serves to reinforce desired outputs (reward) and to inhibit undesired ones (punishment) (Russell & Norvig 2003: 650, cf. Chapter 2.2.1).
- *Speedup learning*. Learning how to perform a previously mastered task more efficiently (faster) from experience, by remembering and analyzing previous performances of the task (Tadepalli & Natarjan 1996; Dietterich 2003; Fern 2007).
- *Analogy*. Learning in which knowledge from one domain is transferred to another based on the observation that there are certain structural similarities between the two domains.

- *Clustering*. Learning that assigns the items of a data set to subsets (clusters) whose members share certain properties, such as similarity or proximity according to some distance measure. Clustering is a form of unsupervised, inductive learning ([INT 121]).
- *Decision tree induction*. Learning that induces the simplest decision tree (cf. Section 12.2.1) which represents the relevant generalizations from the example data and allows the correct classification of both old and new examples (Cawsey 1998: 150–153; Mooney 2003: 379ff.).
- *Discovery*. Learning without supervision or a pre-defined goal, that combines deductive and inductive methods.
- *Genetic algorithms*. Learning that is based on the ideas of "natural selection" and "survival of the fittest" from evolutionary biology. The starting point for a genetic algorithm is a population of candidate solutions to a problem. In each of a series of generations, the 'fittest' (best) solutions are selected (based on an evaluation of their performance), mutated, and combined to produce 'offspring' (i.e. new solutions) that form a new, more evolved population with which the process is repeated (Cawsey 1998: 153–156, [INT 122]).
- *Artificial neural networks (ANNs)*. Learning that involves training networks which link a large number of simple processing units by means of weighted connections. All units in such a network operate in parallel without a central control mechanism. The learning process modifies the weights on the connections and thus influences the future operation of the network (Cawsey 1998: 156–163, cf. Chapter 11.1.1.2 and Section 12.2.1).

The machine learning techniques chosen are preferably compatible with and complementary to the efforts of the knowledge engineer (cf. Section 12.1.1.1). Whether hand-coded or derived automatically, the acquired knowledge then has to be refined, debugged, and reorganized in several successive cycles in order to improve the efficiency and accuracy of the agent or program.

12.1.2 Knowledge Representation

The majority of research in the history of artificial intelligence has been based on the *(physical) symbol system hypothesis* (Newell & Simon 1976; Copeland 1993: Chapter 4; Simon 1996), according to which

> *[a] physical symbol system [(e.g. a computer or a human being)] has the necessary and sufficient means for general intelligent action. (Newell & Simon 1976: 116).*

This hypothesis implies that artificial agents and other computer-based systems can achieve intelligent behavior by manipulating structures that represent items of knowledge in terms of combinations of *symbols* standing for entities and relationships between these entities in the domain (Cawsey 1998: 9; Russell & Norvig 2003: 18).[172] For example, the symbol `berlin` may be used to denote a particular city, the symbol `germany` to label a particular country, and the symbol `Capital(x,y)` to express the relationship between a city `x` and a country `y`, that the former is the capital of the latter. Hence, the *symbol structure* `Capital(berlin,germany)` (in

[172] See Chapter 14.1.2 for a discussion of a more recent, alternative approach to the computation of intelligent behavior that relies less on the manipulation of internal symbolic representations and more on the interaction between an agent and its environment, combining simple behaviors bottom-up into more complex behaviors.

predicate-argument notation, cf. Section 12.1.2.1) captures the fact that the capital of Germany is Berlin.

The symbol structures encoding the knowledge that an agent or system has about its domain are stored in its knowledge base or domain model. To facilitate encoding domain-related knowledge, various *knowledge representation languages (KRLs)* have been developed (Cawsey 1998: Chapter 2; Russell & Norvig 2003: Chapter 10; Brachman & Levesque 2004). In general, a KRL should provide a formalism in which facts of adequate complexity can be represented in a way that is unambiguous and accurate but also comprehensible, and furthermore facilitates the derivation of representations capturing new facts from representations of existing knowledge (Cawsey 1998: 10). To achieve this, a knowledge representation language requires (p. 12):

- *Representational adequacy*. It should be possible to represent the whole range of knowledge required by the application in a structured way that facilitates reasoning with the knowledge.
- *Inferential adequacy*. The language should support inferences from represented knowledge to derive new facts.
- *Inferential efficiency*. Making inferences from existing knowledge should require as little time and as few computational resources as possible.
- *Clear syntax and semantics*. Every expression in the language should have a well-defined structure and meaning.
- *Naturalness*. It should be easy for humans to understand and use the language.

Four[173] major families of symbolic knowledge representation languages have been used in the field of artificial intelligence: logic, semantic networks, frames and scripts, and rules. These formalisms are discussed in the subsections below. A final subsection explores the possibility of creating knowledge representations outside an agent by adding semantic annotations to the agent's environment.

12.1.2.1 Logic

Formal logic, especially predicate logic and its derivatives, was one of the first schemes for the representation of knowledge used in the field of artificial intelligence (McCarthy 1958), and its role as the most important knowledge representation language for intelligent systems continues (Cawsey 1998: 20). Logic is fundamental in the sense that representations in other formalisms (semantic networks, frames, scripts, and rules) can be translated into equivalent logical formulae with relative ease (p. 37).

Predicate (or *first-order*) *logic* (Cawsey 1998: 20–29; Russell & Norvig 2003: Chapters 8–9; Brachman & Levesque 2004: Chapter 2) represents facts in terms of *predications*, i.e. symbol structures that express a property of a single entity or a relationship among several entities. The property or relationship is encoded by the *predicate* whose *argument(s)* denote the entity or entities involved (Cruse 2000: 19). For example, the predication *Hate(john,milk)* consists of the predicate *Hate(x,y)*, which, in turn, takes two arguments, *john* and *milk*. The internal structure of predications represents the structure of the particular state of affairs in the world which they describe. In particular, the order of arguments in a predication is important: *Hate(john,milk)* and

---

173 It can be argued that frames and scripts are essentially a variant of semantic networks (Cawsey 1998: 16), so there are only three major approaches to the symbolic representation of knowledge. However, for expository reasons, frames and scripts are discussed separately from semantic networks in this chapter.

Table 18. Some facts and their representation in predicate logic.

Fact	Representation in Predicate Logic	Description
John is human.	Human(john)	“The entity ‘john’ is human.”
Humans are mortal.	$\forall$x(Human(x) $\rightarrow$ Mortal(x))	“For all x, it holds that if x is human then x is mortal.”
Robins are birds.	$\forall$x(Robin(x) $\rightarrow$ Bird(x))	“For all x, it holds that if x is a robin then x is a bird.”
Birds have wings.	$\forall$x(Bird(x) $\rightarrow$ Have_wings(x))	“For all x, it holds that if x is a bird then x has wings.”
Berlin is the capital of Germany.	Capital(berlin,germany)	“The entity ‘berlin’ is the capital of the entity ‘germany’.”
The yellow box is between the red box and the green box.	Between(yellow_box,red_box,green_box)	“The entity ‘yellow_box’ is between the entity ‘red_box’ and the entity ‘green_box’.”
Vampires do not exist.	$\neg$x(Vampire(x))	“There is no x for which it holds that x is a vampire.”
All books are written by at least one author.	$\forall$x$\exists$y(Book(x) $\rightarrow$ (Author(y) $\wedge$ Write(y,x)))	“For all x, it holds that there is a y such that if x is a book then y is an author and y wrote x.”

?Hate(milk,john) make two different statements about the world, the latter of which is doubtful (indicated by the ‘?’) because it contradicts common-sense knowledge.

Depending on the number of arguments involved, a distinction is made between *one-place*, *two-place*, *three-place*, and *n-place* predicates, which take one, two, three, and n arguments, respectively (Cruse 2000: 20). Expressions in predicate logic may also use *quantifiers*, i.e. operators that allow the formulation of statements about all, some, or none of the (anonymous) entities (represented by *variables x*, *y*, etc.) in a given collection (Jurafsky & Martin 2000: 517ff.; Russell & Norvig 2003: 249–253). The set of quantifiers includes the *universal quantifier* $\forall$ (implies *universality*; the expression affected by the quantifier holds for all entities), the *existential quantifier* $\exists$ (implies *existence*; the expression holds for at least one entity), and the *negative quantifier* $\neg$ (implies *non-existence*; there is no entity for which the expression holds). Furthermore, *logical connectives* are used to combine simple predications into expressions describing more complex states of affairs. The set of connectives comprises $\wedge$ (*conjunction*, “and”), $\vee$ (*disjunction*, “or”), $\leftrightarrow$ (*equivalence*, “if and only if”), $\rightarrow$ (*implication*, “if .. then”), and $\neg$ (*negation*, “not”) (Cawsey 1998: 21; Russell & Norvig 2003: 204ff.). Several facts and their representations in predicate logic are listed in Table 18.

Predicate logic provides a *formal language*, based on the notion of *truth* (i.e. statements can either be true or false), in which facts about the world can be expressed in terms of statements about entities and the relationships between them, which can be generalized to entire collections of entities. The syntax and semantics of this language are well-understood, so the structure and meaning of statements formulated in predicate logic are clear (Antoniou & van Harmelen 2008: 157). Proof systems based on predicate logic are *sound* (i.e. everything that is concluded follows from the premises) and *complete* (i.e. everything that follows from the premises can be derived) (pp. 157f.). Automated reasoners can use them to reveal implicit knowledge by drawing conclusions (*inferences*) from the facts that are explicitly represented in the knowledge base (p. 13). One of the most important inference rules in logic is *modus ponens* (if-then reasoning) (Cawsey 1998: 22; Jurafsky & Martin 2000: 520; Russell & Norvig 2003: 211). It is defined as follows (p and q stand for logical formulae; $\Rightarrow$ means 'yields' or 'proves'):

$$(p \wedge p \rightarrow q) \Rightarrow q$$

$$(\text{Human(john)} \wedge \forall \text{x(Human(x)} \rightarrow \text{Mortal(x))} \Rightarrow \text{Mortal(john)}$$

According to modus ponens, the formula on the right of $\Rightarrow$ can be derived from the formula in brackets on the left in the following way: if the knowledge base has the antecedent (p) of an implication rule ($p \rightarrow q$), it is possible to infer the consequent of the rule (q) (Jurafsky & Martin 2000: 520). In the example, given the knowledge that John is human and that all humans are mortal, it can be inferred that John is also mortal. There are two practical uses of modus ponens (pp. 520f., cf. Section 12.1.2.4):

- *Forward chaining*. The reasoner automatically adds more facts to the knowledge base by applying all implication rules matching a new fact.
- *Backward chaining*. The reasoner proves query formulae not contained in the knowledge base by showing the truth of the antecedent of any implication rule whose consequent matches the query formula.

Apart from inferring implicit knowledge, automated proof systems based on logic are also useful for identifying inconsistencies in the knowledge base (e.g. the inconsistency of having **(Male(john) ∧ Female(john))*), and for helping intelligent agents in decision making and action selection (Antoniou & van Harmelen 2008: 14). The possibility for an agent to explain chains of reasoning that underlie its action or answer (by going backwards through the sequence of inference steps) makes the agent's reasoning processes transparent to the user (p. 14), which is an important aspect of several instructional interactions of pedagogical agents (cf. Section 12.3) and a highly desirable feature in other domains that involve the user delegating tasks to and thus having to trust the agent, such as electronic commerce (cf. Chapter 3.2.2).

Predicate logic is sufficiently expressive to allow the representation of a wide range of reasonably complex declarative knowledge. In addition, it provides a well-defined syntax and semantics, as well as sound and complete proof systems that facilitate efficient inferences from existing knowledge using automated reasoning tools (Cawsey 1998: 12 + 20; Russell & Norvig 2003: 240). While other logics are more expressive than predicate logic, computing inferences from their representations is generally more expensive and in certain cases even impossible (Antoniou & van Harmelen 2008: 14).

However, predicate logic cannot be recommended as a knowledge representation language without some reservations. First, if the application involves only simple facts and inferences, the expressive and inferential power of a logic-based system may be overkill (Cawsey 1998: 12). A simpler language, for example one based on frames (cf. Section 12.1.2.3) or rules (cf. Section 12.1.2.4), may be more than sufficient for these applications (p. 13). Another problem arises with respect to the possibilities of knowledge organization. Logic-based (as well as rule-based) knowledge representations are *flat*, i.e. they consist of unstructured collections of self-contained statements that can be interpreted without referring to other statements. As a result, collecting all available knowledge about a particular entity or class of entities becomes difficult because it may be scattered among numerous statements that have no obvious relation to each other (Brachman & Levesque 2004: 135). In contrast, semantic networks (cf. Section 12.1.2.2), frames, and scripts (cf. Section 12.1.2.3) provide structures that organize knowledge about entities, classes, or activities in a way that clearly indicates which items of a representation relate to which entity, class, or activity.

A third issue concerns limitations of the expressive power of predicate logic. There are certain types of knowledge that are difficult to represent in predicate logic (Cawsey 1998: 27; Brachman & Levesque 2004: 44):

- *Statistical facts*. Statements that hold for all, none, or at least one member of a set are straightforward in predicate logic; however, it is much harder to specify subsets that can be quantified more or less precisely (e.g. "half of the students," "most of the instructors," and "almost none of the laboratory staff").
- *Uncertainty*. In complex, dynamic, or inaccessible environments, agents will always operate under uncertainty, i.e. lack of complete knowledge (Russell & Norvig 2003: 462), due to both *laziness* (rules without exceptions are too difficult to develop and apply) and theoretical (there is no complete theory for the application domain) or practical *ignorance* (not all pertinent information has been or can be obtained) (pp. 463f.). As a result, the agent can only have a certain *degree of belief* in the truth of statements (e.g. "Fuel prices will rise next month"), which is expressed as a probability between 0 and 1 (p. 464). Degree of belief cannot be captured by predicate logic.
- *Fuzzy knowledge*. Vague descriptions and the extent to which they apply to entities in the domain constitute another problem for predicate logic (Russell & Norvig 2003: 526). For example, the predication *Tall(john)* states that a certain individual with the name John is 'tall.' The problem is that the word 'tall' used as the label of the predicate does not define a clearly demarcated category of sizes for which the predicate is true. In other words, tallness is a matter of degree, and so is the truth of the statement *Tall(john)* (p. 526). In predicate logic, however, this statement can only be either true or false, i.e. either John is a member of the set of tall people or not. *Fuzzy logic* (Zadeh 1965) is a theory of multi-valued logic (as opposed to binary predicate logic) that provides a way to reason with logical statements like *Tall(john)*, which denote membership in a vaguely defined (*fuzzy*) set of entities. In this framework, *Tall(x)* would be regarded as a *fuzzy predicate*, and *Tall(john)* would be assigned a number between 0 and 1 that represents its *degree of truth* (Russell & Norvig 2003: 526f., [INT 123]).
- *Default and prototypical knowledge*. Certain facts are normally (assumed to be) true, unless (or until) there is contradictory evidence. Examples include "Exchange students speak English," "Computers have one central processing unit," and "People applying for Job X have Qualification Y." It is difficult to create statements in predicate logic which are initially held to be true but may turn out to be false, or which have exceptions.
- *Propositional attitudes*. Predicate logic does not provide adequate means for agents to represent and reason about what they (or other agents) *believe*, *know*, or *want*, i.e. their mental attitudes (or those of other agents) toward a statement (or proposition) (Russell & Norvig 2003: 341). The reality of certain propositional attitudes for an agent does not necessarily mean that the outside world concurs with them. For example, John may believe that pigs can fly, but they cannot; Mary may know that London is the capital of Australia, but it is not; or Paul may not want Linda to know that he loves her, but she does.
- *Time*. There is no straightforward way to represent time using predicate logic. Expressions in predicate logic generally do not provide temporal information, so *Write(peter,letter)*, for example, could mean "Peter wrote a letter," "Peter is writing a letter," or "Peter will write a letter." One way to address this problem is to add temporal variables and temporal predicates to the representation. Temporal variables capture the time interval and the end point of that interval that are associated with a given event. Temporal predicates like *Precedes* indicate how the end point of the interval is related to the current time (represented by the constant

Now). For events in the past, *Precedes(e,Now)* locates the end point before the current time, whereas *Precedes(Now,e)* states that the current time precedes the end point of a future event. For present events, the current time lies within the time interval of the event (Jurafsky & Martin 2000: 528). The first logical system for the representation of time was *temporal logic*, which is based on the assumption that logical statements are true at particular points or intervals of time, which are ordered (Russell & Norvig 2003: 244). In this framework, a set of modal operators is applied to logical statements to indicate, for example, that a statement (proposition) p will be true at *all* times in the future ($\Box p$) or at *some* time in the future ($\Diamond p$) (p. 364).

- *Change*. Another point of criticism concerns the ‘static’ nature of predicate logic. In a given interpretation, statements in predicate logic are either true or false and keep their truth value, regardless of what changes occur in the world or in an individual’s beliefs about the world (Brachman & Levesque 2004: 285). However, agents need to consider both sorts of changes: they have to reason about a changing environment, and they may have to revise (change) some of their beliefs about the environment. *Situation calculus* (Russell & Norvig 2003: 328–340; Brachman & Levesque 2004: Chapter 14) is a dialect of predicate logic that allows to represent an agent’s beliefs about a changing world (Brachman & Levesque 2004: 286). In this framework, change occurs when an action takes an agent from a given static situation to a new situation. For example, if a robot picks up a bucket of water, that action takes it from one situation (not holding the bucket) to another (holding the bucket) (p. 286). Three basic elements are involved in expressions used in situation calculus (Russell & Norvig 2003: 329; Brachman & Levesque 2004: 286f.; [INT 124]):
  - *Actions* that can be carried out in the environment (e.g. *Pickup(r,x)* “robot *r* picking up object *x*”);
  - *Fluents*, i.e. predicates and functions with (possibly) different values across situations, which represent what is true in a given situation (e.g. *Holding(r,x,s)* “robot *r* holding object *x* in situation *s*”);
  - *Situations* consisting of histories of situations, where each situation results from applying an action to the previous situation. The initial situation to which the first action is applied is called S_0. A function *Do(a,s)* maps from an action *a* performed in situation *s* to a certain resulting situation. For example, $Do(Pickup(r,b_2),Do(Pickup(r,b_1),S_0))$ is a situation term corresponding to the situation resulting from robot *r* first picking up object b_1 in the initial situation S_0 and then picking up object b_2 in the following situation.

12.1.2.2 Semantic Networks

The term *semantic network* (Collins & Quillian 1969; Cawsey 1998: 14f.; Russell & Norvig 2003: 350ff.; Brachman & Levesque 2004: Chapter 10[174]) refers to a group of knowledge representation formalisms which have in common that they represent knowledge as a network of labeled *nodes* and *arcs* (or *links*). The nodes in the network stand for concepts (entities, events, and states) of the domain. Arcs represent the relations between concept nodes (Cawsey 1998: 14). There can be various such relationships, like the taxonomic `IS-A` relation (e.g. `CANARY IS-A BIRD`), the property-assigning `HAS` relation (e.g. `ROBIN HAS WINGS`), and the

[174] Brachman and Levesque used the term “inheritance network” (Brachman & Levesque 2004: 188), which emphasizes the possibility of inheritance reasoning enabled by these networks (see below).

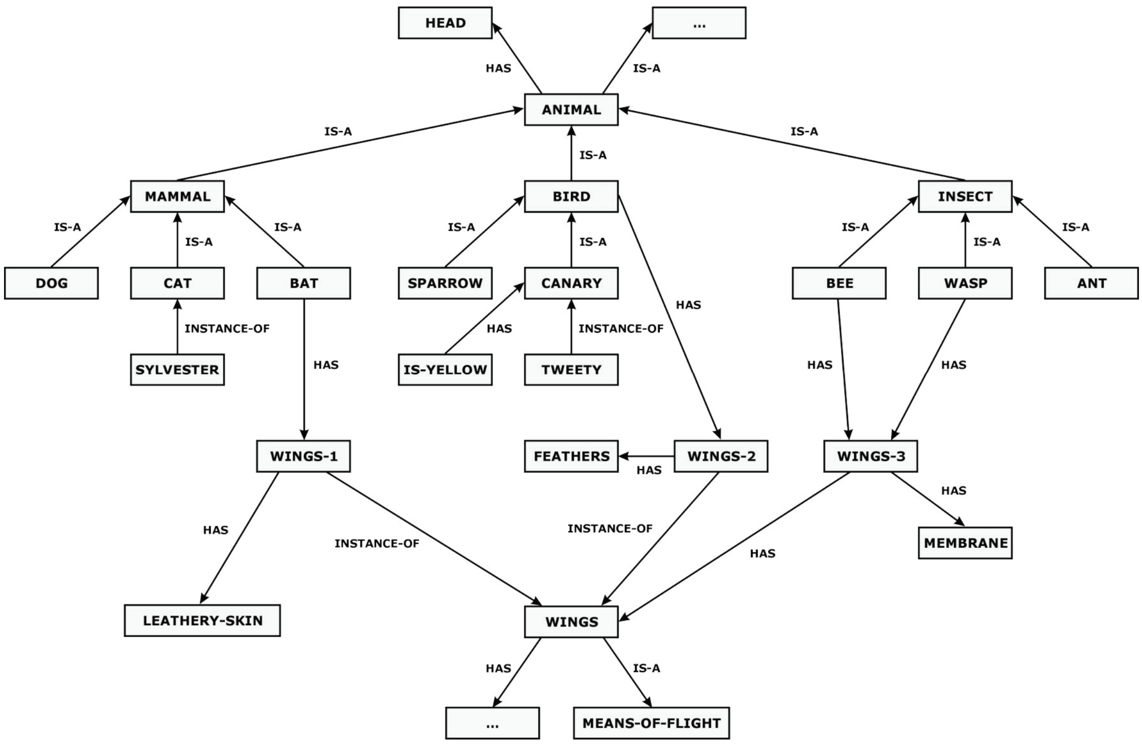

Figure 55. Part of a semantic network showing taxonomic (IS-A), property (HAS) and instance (INSTANCE-OF) relationships.

classifying `INSTANCE-OF` relation (e.g. `TWEETY` is an `INSTANCE-OF` the category (class) `CANARY`). The direction of a relation is often explicitly shown in the network by means of arrows but may also be given by the underlying theory. Figure 55 provides an example of a partial semantic network of animals, their attributes, and taxonomic and instance relationships.

Semantic networks provide a simple and intuitive formalism for capturing knowledge about entities in a given domain and the relationships between them. The graphical network notation makes the organization of the domain knowledge transparent for human readers (Cawsey 1998: 15). Another major feature of semantic networks is that their formalism minimizes redundancy in the knowledge representation and facilitates inferences from the represented knowledge. The `IS-A` links between nodes in a semantic network define a *property inheritance hierarchy* among the corresponding domain concepts. By default, lower nodes in this hierarchy inherit properties associated with higher nodes. The inherited properties can be kept or redefined, and new ones can be added. Given this mechanism, it is possible to infer from the semantic network in Figure 55 that, for example, `TWEETY` has the attribute `IS-YELLOW` and possesses a particular kind of wings (`WINGS-2`) and a `HEAD`, by following the link from `TWEETY` to the class (node) of which `TWEETY` is an instance (`CANARY`) and then moving up the `IS-A` hierarchy from there, collecting the properties linked to the nodes in the hierarchy along the way. With property inheritance, both overloading individual concepts with properties and redundant specification of properties across the network can be avoided. However, despite their elegance, semantic networks have several shortcomings:

- Their complexity quickly becomes overwhelming with the rapidly growing number of nodes and links required to represent more, and more complex knowledge (Handke 1995: 96f.).
- The interpretation (semantics) of the nodes and links in semantic networks is application- and theory-dependent. Hence, the validity of inferences made from a given network is not

guaranteed, because its representation might be misinterpreted (Cawsey 1998: 14f.).[175] In contrast, the semantics of logical expressions is clear, and inferences from them are assured to be valid (cf. Section 12.1.2.1).

- Semantic networks only support inheritance reasoning to infer the properties of subclasses from the properties of parent classes in the hierarchy (Cawsey 1998: 37).[176]
- In comparison to predicate logic, it is difficult to express negation (something is not true), disjunction ((either) A or B is true), and quantification (something is true of all or some (at least one) of the entities in a set) (Cawsey 1998: 20).

12.1.2.3 Frames and Scripts

When people encounter a new situation, their previous experience supplies them with a set of expectations, or predictions, about what they will find and how they should behave in that situation. Frames and scripts are two related representation languages for encoding knowledge about the entities and events that are (stereo-) typically involved in particular situations.

Frames (Minsky 1975; Cawsey 1998: 16–20; Nebel 2001; Brachman & Levesque 2004: Chapter 8) organize stereotyped knowledge about an entity, event, or situation in a container structure (cf. Figure 56) consisting of attribute-value pairs (*slots* and *fillers*), which resembles the record (or 'struct') data structures used in programming languages (Cawsey 1998: 16). Slots of superordinate frames are inherited through a hierarchy by subordinate frames, as well as by particular instances of frames (Nebel 2001: 324). A given frame can have several parent frames (*multiple inheritance*) (Cawsey 1998: 18). In general, property inheritance in frame hierarchies is *defeasible*, i.e. an inherited slot filler provides a *default value* that is kept only if no filler has been provided or can be computed. Hence, slot fillers may represent typical values that can be overridden by a different value in both instances and specializations of a generic frame (Cawsey 1998: 17; Brachman & Levesque 2004: 139). Furthermore, slots can have procedures attached to them, i.e. programmed routines which derive missing values (`IF-NEEDED`) or perform an action if the value of a slot is deleted (`IF-REMOVED`) or added/updated (`IF-ADDED`) (Cawsey 1998: 19; Brachman & Levesque 2004: 138ff.).

Like semantic networks, frames represent (categories of) objects, their properties, and their relationships,[177] but in a more compact and encapsulated ('object-oriented') format (Nebel 2001: 324).[178] Property inheritance provides them with a mechanism to infer knowledge that is

---

[175] In general, a precise formulation of the semantics of a knowledge representation language is essential for knowing the exact meaning of each expression written in the language as well as the range of sound inferences from known facts (Cawsey 1998: 15).

[176] It should be noted that the restriction to inferences based on property inheritance is not necessarily a weakness of semantic networks (or frames, which are restricted in the same way). In fact, if the application involves the representation of declarative knowledge about related concepts or objects that can be arranged in a taxonomic hierarchy, and only requires a single type of inference, frames and semantic networks may be more adequate than more complex formalisms, such as logic (Cawsey 1998: 37).

[177] Both frames and semantic networks can be translated into the other representation with relative ease. The nodes in the semantic network become frames for classes and instances, their outgoing links become slots in those frames, and the nodes at the other end of the links fill the slots in the frames (Cawsey 1998: 16).

[178] Frame systems share terminology and ideas with the *object-oriented programming (OOP)* paradigm, which was concurrently developed and influenced by them (Cawsey 1998: 16; Nebel 2001: 324).

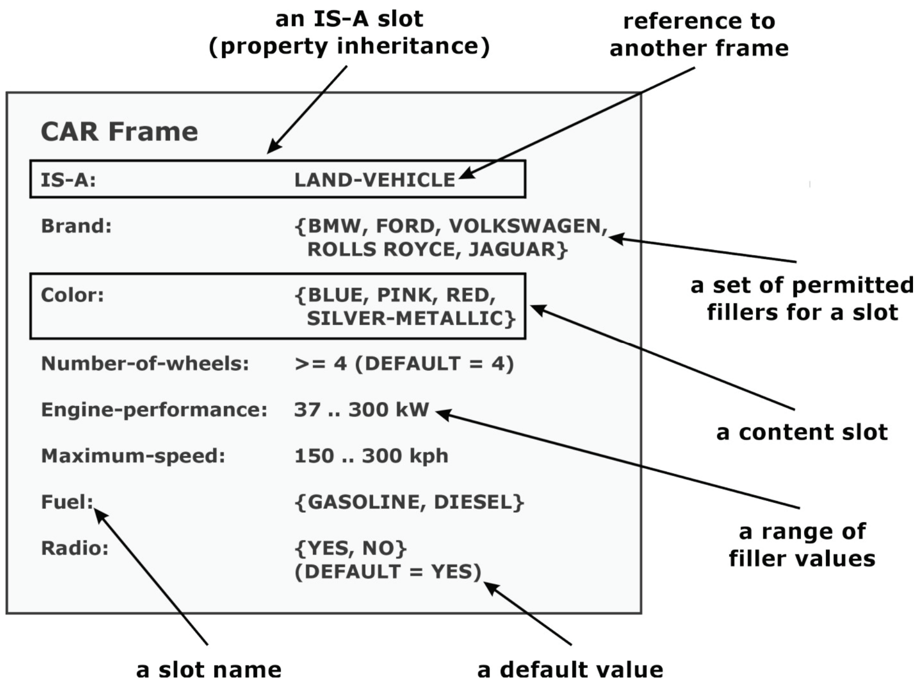

Figure 56. The components of a frame.

not explicitly stated in a given frame but inherited from parent frames in the hierarchy (Cawsey 1998: 20). In fact, frame-based reasoning involves instantiating generic frames, supplying values for some of the slots in the instance frames, and computing other slot values by way of property inheritance or execution of procedures attached to slots (Brachman & Levesque 2004: 138f.). Standard expectations can be encoded as default values, and missing or changing values can trigger appropriate computational procedures.

Frames (and semantic networks) are appropriate for representing declarative knowledge that can be organized in terms of taxonomic hierarchies. However, things like negation, disjunction, and quantification can be better expressed using predicate logic (cf. Section 12.1.2.1). The only type of inference supported is enabled by property inheritance: properties are passed down from higher levels down to lower levels in the hierarchy (Cawsey 1998: 20). However, this may be sufficient for quite a number of applications. Furthermore, the frame mechanism itself gives rise to several issues:

- Is there a universal set of slots that can be used across domains? What should be included in this set? Is the set finite?
- Slot names and fillers are generally arbitrary labels whose meanings are not formally and clearly defined. *Terminological* or *description logics (DL)* allow to specify the semantics of the expressions and relations used in frame-based languages in terms of a structurally and formally well-understood logic (Cawsey 1998: 27f.; Russell & Norvig 2003: 353f.; Brachman & Levesque 2004: Chapter 9). Having a formal specification of the semantics in a

Like frames, OOP involves classes, instances (objects), properties (attributes or fields), procedures (methods), and inheritance; however, instead of knowledge representation, OOP is concerned with writing computer programs that consist of encapsulated, cooperating objects with different roles and responsibilities that work by exchanging messages with other objects and internally processing data (Cawsey 1998: 16; Booch et al. 2007; [INT 125]).

terminological logic helps to ensure the soundness of all inferences obtained from the frame structure (Cawsey 1998: 27).

- When default values of slots can be overridden in an unconstrained fashion, the resulting frame structure may be at odds with human understanding of the domain or the world in general (Brachman & Levesque 2004: 140). For example, it is possible to create an instance or specialization of the `ELEPHANT` frame in which the default value of the `Habitat` slot is overwritten with `TREE` (a reference to the `TREE` frame), making that (kind of) elephant one that lives in trees, which is quite difficult to imagine given the size, weight, and extremities of these animals (Brachman 1985: 81).

Scripts (Schank & Abelson 1977) are basically a kind of frame that encodes stereotyped event sequences representing socially ritualized human activities including visiting a restaurant, buying a car, and opening a bank account (Dyer et al. 1992: 1443f.). An example of a script for eating at a restaurant is provided in Figure 57. Each event in a script is described in terms of a *conceptual dependency (CD-) structure* (Schank 1972) defining a network of relationships (dependencies) between a primitive action, an actor performing the action, an object affected by the action, and a direction specifying source and goal of the action (Handke 1995: 103). Actions in a CD-structure can be 'stand-alone' *basic acts* (`ATRANS`, `PTRANS`, `MTRANS`, `MBUILD`, `PROPEL`, and `INGEST`) or 'supporting' *instrumental acts* (`GRASP`, `MOVE`, `SPEAK`, `ATTEND`, and `EXPEL`) (p. 104).

Scripts have had a number of successful applications in natural language processing (Dyer et al. 1992: 1449–1456). In particular, they have been used to understand simple stories, as in the SAM (Script-Applier Mechanism) system (Cullingford 1978).[179] But relying only on scripts to accomplish story understanding and other tasks raises several issues (Barr & Feigenbaum 1981: 308):

- The number of possible stories is overwhelming. Hence, story understanding cannot work by simply matching scripts to stories, even if scripts are defined as generally as possible and a computer knows a large number of them.
- Scripts describe a complete sequence of events. But more often than not, everyday activities (or accounts of them) remain unfinished, or they are interrupted by other, unrelated events. How can a script-based system reliably detect transitions and interactions between scripts when these are not clearly marked?
- The execution of scripts can be hindered by obstacles, like, for example, the application of the `LIBRARY` script: "I went to the library for a book I needed for my Ph.D. research. But the library was closed."
- Stories may provide insufficient information for a computer to activate a script, but people can still understand them by making bridging inferences (cf. Chapter 11.1.3.3), e.g. "Paul wanted to marry Linda. He picked up a gun and went to her father."
- The CD-format used for the event descriptions in scripts may be too restricted for realistic applications. Specific shortcomings of the original CD-theory concern the handling of

[179] The ability to understand stories may be much more important to the operation of intelligent systems than it seems. Narrative psychology (cf. Chapter 7.7) regards stories, or narratives, as fundamental to making sense of the behavior of intentional beings. So by assimilating the observed actions of others into narrative structures, intelligent systems can see how these actions fit into a larger scheme, what they are eventually aiming at, what the other might do next (i.e. how the story might continue), and so on.

EAT-AT-RESTAURANT Script

Goal:	**eat-food**
Props:	**tables, chairs, menu, bill, money, selection-of-food**
Roles:	**guest, waiter, cook, owner**
Preconditions:	**guest-is-hungry, guest-has-money**
Results:	**guest-has-less money, owner-has-more money, guest-is-not-hungry**

Scene 1: Entering

guest PTRANS from unknown-location to restaurant
guest ATTEND eyes-of-guest to tables
guest MBUILD free-table
guest PTRANS guest to free-table

Scene 2: Ordering

waiter PTRANS waiter to guest's-table
waiter ATRANS menu to guest
guest MTRANS menu to intermediate-memory
guest MBUILD selection-of-food
guest MTRANS selection-of-food to waiter
guest ATRANS menu to waiter
waiter PTRANS waiter to cook
waiter MTRANS selection-of-food to cook
cook DO (PREPARE-FOOD Script)

Scene 3: Eating

waiter PTRANS waiter to guest's-table
waiter ATRANS selection-of-food to guest
waiter PTRANS waiter to unknown-location
guest INGEST selection-of-food

Scene 4: Paying

guest MTRANS to waiter
waiter DO (WRITING-THE-BILL Script)
waiter PTRANS waiter to table
waiter ATRANS bill to guest
guest ATRANS money to waiter
guest PTRANS guest from restaurant to unknown-location

Figure 57. A script for eating at a restaurant, subdivided into four scenes. Adapted from Schank and Abelson (1977: 43f.) and Schmitz (1992: 69). The events in each scene are specified as pseudo-natural language descriptions around a primitive act (e.g. DO, MBUILD, PTRANS) from conceptual dependency theory (written in capital letters).

quantification and metaphor as well as the theory's focus on action verbs, whose meanings were reduced to their physical realizations (e.g. kiss = "MOVE lips to lips") (Barr & Feigenbaum 1981: 214). Later extensions to CD-theory included mechanisms for the representation of social and interpersonal activities (Schank & Carbonell 1978).

One alternative to having a set of static, pre-fabricated scripts for different kinds of scenarios is to dynamically construct a *plan* (or alternative plans) that account(s) for a given situation by integrating the observed actions of an actor into a sequence of actions that, if carried out, would achieve the identified goals and subgoals of that actor (Allen 1994: 480). Plan-based story understanding then involves establishing the actor's goals and their associated instrumental subgoals from the story and building plans that achieve goals and subgoals while incorporating the actor's observed actions (Schank & Abelson 1977; Barr & Feigenbaum 1981: 309f.). Unlike scripts, plan-based approaches allow for the dynamic construction (for different situations and from individual observed actions), incompleteness, and as-needed revision of representations. See Chapter 12.1.3.3 for a discussion of planning.

12.1.2.4 Rule-Based Systems

Rule-based systems (also called *production systems*) (Cawsey 1998: 29–36; Simon 2001; Russell & Norvig 2003: 280–295; Brachman & Levesque 2004: Chapter 7) are a popular way to represent and reason with knowledge in expert systems (Cawsey 1998: 13, cf. Footnote 171). The knowledge of a rule-based system consists of a set of *IF-THEN (condition-action) rules* (often called *production rules* or *productions*), which define what to do or conclude given certain conditions (Cawsey 1998: 29). Figure 58 shows the main components of a rule-based system (Cawsey 1998: 29f.; Brachman & Levesque 2004: 118f.):

- A constantly updated database of assertions (*facts*) about the domain. Another term for this component is *working memory*.
- A collection of *rules* of the format "IF conditions THEN actions," which specify a set of conditions (the *antecedent*) that have to be satisfied by the current state of the working memory for the actions in the body of the rule (the *consequent*) to be considered for execution. If this happens, the actions performed change the working memory.
- A *control scheme (interpreter)*, which identifies applicable rules, selects a subset of them for execution (using *conflict resolution* strategies), and *fires* the selected rules, i.e. performs their action part, which causes the working memory to be updated.

Each rule is a self-contained piece of knowledge, whose application is controlled by the interpreter, guided by the contents of the working memory (Cawsey 1998: 29). While condition-action rules are reminiscent of implications in predicate logic (cf. Section 12.1.2.1), rule-based languages are more procedural than declarative in nature (p. 13), i.e. rather than making statements about the world which are true or false, as in logic, production rules state what actions to take or what conclusions to draw in a given situation; in other words, they encode more of the 'know-how' of a domain than of its 'know-what' (cf. Table 17) (p. 37).

Both rules and facts may contain variables. In facts, these variables carry specific values, as in the two facts `F1` and `F2` below; in rules, they form *patterns* that the interpreter matches against the instantiated facts in the working memory (cf. `R1` and `R2`) (Cawsey 1998: 32):

```
F1: number_of_terms(peter,2)
F2: passed_introlin(peter)

R1: IF number_of_terms(S,N) < 5 THEN
       ADD undergraduate_student(S)
R2: IF undergraduate_student(S) AND passed_introlin(S) THEN
       ADD cleared_for_ps(S)
```

In rule `R1`, the pattern `number_of_terms(S,N)` matches the fact `F1`, the condition is satisfied, and the rule fires, adding the new fact `F3: undergraduate_student(peter)` to the working memory. The facts `F2` and `F3` match the patterns `passed_introlin(S)` and `undergraduate_student(S)`, respectively, so rule `R2` can fire, adding the new fact `F4:`

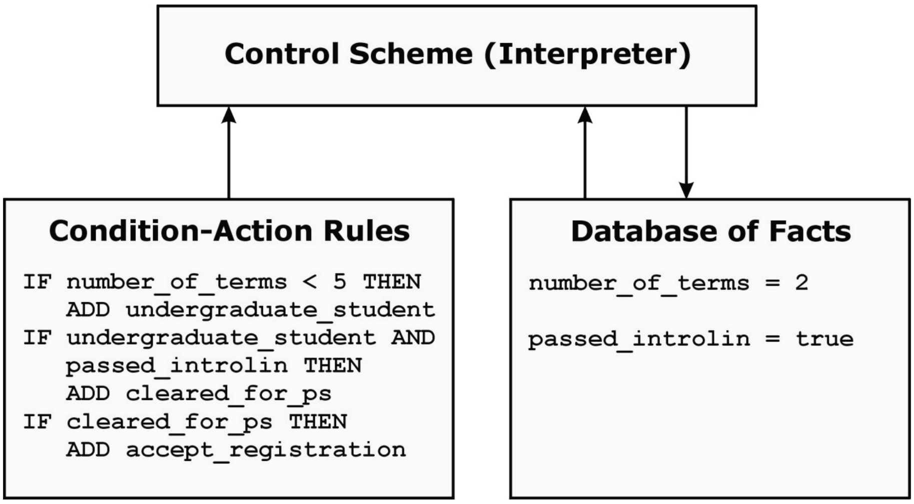

Figure 58. The architecture of a rule-based system. Adapted from Cawsey (1998: 29, Figure 2.6).

`cleared_for_ps(peter)`. It is also possible to match goals whose variables are instantiated with specific values, e.g. `G1: accept_registration(peter)`, against the parameterized conclusions (actions) of rules (Cawsey 1998: 33), as in rule `R3` below:

```
R3: IF cleared_for_ps(S) THEN
       ADD accept_registration(S)
```

Two major classes of algorithms are used in production systems to reason with the rules: data-driven (or bottom-up) forward chaining and goal-driven (or top-down) backward chaining. *Forward chaining* starts with the available facts in the working memory and repeatedly fires applicable rules to derive new facts (or perform specified actions) based on the evolving state of the fact database. A given rule becomes available for execution if the working memory contains facts that match its antecedent. A rule that is fired typically modifies the working memory by adding or deleting facts, but its action part may also invoke procedures that perform certain operations, such as animating an on-screen character (Cawsey 1998: 30). The basic algorithm of forward chaining is a selection-execute cycle (pp. 30f.). Conclusions derived during forward chaining may become invalid because of changes to the facts that serve as their justifications. In consequence, these obsolete facts have to be retracted, i.e. removed from the working memory (Russell & Norvig 2003: 360). If additional facts have been inferred from the obsolete facts, those inferred facts cannot simply be deleted as well because the invalidated facts may not be their only justifications (p. 360). *Truth maintenance systems (TMSs)* efficiently perform these kinds of updates, resolving conflicts and ensuring the consistency of the fact database (Cawsey 1998: 32; Russell & Norvig 2003: 360ff.).

While forward chaining works from existing facts to new conclusions, backward chaining takes the reverse direction (from conclusions to facts) (Brachman & Levesque 2004: 92f.). The starting point is an initial hypothesis (or goal) to be proved. If that goal matches one of the current facts, nothing has to be done. Otherwise, a rule is selected whose consequent matches the goal. The antecedent conditions of that rule become new subgoals that the system tries to prove by recursively applying the same procedure (Cawsey 1998: 34). The initial hypothesis has been proved if the algorithm has managed to prove all the subgoals and has found a chain of rules that link the hypothesis to facts in working memory which provide the basis for the proof (Cawsey 1998: 34f.; Russell & Norvig 2003: 220).

Forward chaining enables agents to perform *data-driven* reasoning. Given incoming percepts from their environment, agents can fire applicable rules to draw conclusions from those inputs, without the need for a particular hypothesis (or goal). However, depending on the rule set and the fact database, a potentially large number of conclusions could be derived this way that are not relevant to the agent's circumstances and tasks, so mechanisms are required that help to avoid flooding the agent's working memory with irrelevant conclusions (Russell & Norvig 2003: 219f.).

The strength of backward chaining is *goal-driven* reasoning, which focuses on just the goal to be proved and does not consider irrelevant facts. In general, backward chaining provides an efficient way for agents to answer specific questions or decide on a course of action to achieve a particular goal (Russell & Norvig 2003: 220). Still, the process can also be computationally expensive if it has to consider many potential hypotheses with many possible ways to prove them (Cawsey 1998: 36). Backward chaining may be combined with forward chaining, with the latter generating facts that are likely to pertain to goals to be proved by the former (Russell & Norvig 2003: 220).

12.1.2.5 The Semantic Web

Instead of trying to rebuild some aspects of a human brain, we are going to build a brain of and for humankind. (Fensel & Musen 2001: 25).

Traditionally, the domain model is developed as a separate module of an intelligent software agent. Because it encodes knowledge about the agent's environment (among other things), it has to be updated whenever that environment changes, which is a problem especially for agents in open, complex, and dynamic environments, such as the World Wide Web, where changes tend to occur frequently and unpredictably. An alternative to developing the domain model exclusively as an internal component of an agent is to represent at least part of this knowledge *externally* by adding annotations to the environment (Doyle & Hayes-Roth 1998; Isbister & Doyle 1999; Doyle 2004), which consist of declarations and procedures attached to objects, tasks, and individuals in different places in the environment (Isbister & Doyle 1999). An agent visiting a particular virtual place reads its embedded annotations to learn about the structure, entities, events, and history of that place, as well as about appropriate behaviors for itself and for others (Doyle & Hayes Roth 1998). The advantage of this approach is that the agent can operate in unknown or changed/changing locations by querying the environment directly rather than by engaging in complex reasoning about the environment. Moreover, the information that the agent has about the environment can be more easily kept up-to-date (Doyle & Hayes Roth 1998).

The concept of annotating resources with semantic information that can be processed by machines is at the core of the emerging *Semantic Web*[180] (Berners-Lee 2000: Chapter 13; Berners-Lee et al. 2001; Fensel et al. 2003; Golbeck et al. 2005; Shadbolt et al. 2006; Ziegler 2007; Antoniou & van Harmelen 2008; [INT 127]), an extension of the current World Wide Web which adds a *semantic layer* of machine-processable annotations to web resources

[180] The vision of the Semantic Web was first formulated by Tim Berners-Lee, the inventor of the World Wide Web, at the first International World Wide Web Conference in Geneva, Switzerland in 1994 (Shadbolt et al. 2006: 96). The realization of the Semantic Web is a collaborative process involving contributors from both academia and industry that is led by the World Wide Web Consortium (W3C) ([INT 127]).

(formulated using shared vocabularies of terms) that provides a common infrastructure for programs, services, devices, agents, and people to consume, produce, and share information on the Web (Berners-Lee et al. 2001; Hendler et al. 2002; Koper 2004: 2; Motta 2006: 88). The goal of building the Semantic Web is for web-based interactions between machines to become as commonplace and straightforward as human-human exchanges mediated by the World Wide Web are today. However, interactions between machines are meant to supplement, not replace those between people (Clark et al. 2004b: 6).

Instead of developing complex software to interpret web content (which is largely in formats more suitable for human understanding than for machine processing), the Semantic Web approach is based on the concept of the *self-describing resource* (Antoniou & van Harmelen 2008: 9f.; [INT 126]). Resources include (web) documents, multimedia elements, people, places, database records, abstract concepts, search queries, etc. – in short, any entity that can be referenced with a unique identifier (Antoniou & van Harmelen 2008: 67f.). Anything uniquely identifiable, whether in virtual worlds or in the real world, qualifies as a resource. On the World Wide Web, users are asked to add *semantic markup* to web resources (Horrocks 2002; Marshall & Shipman 2003), i.e. annotations encoding knowledge about the contents, author, price, etc. of a web page, image, product, and so on in a publicly accessible, machine-processable form. Such "data about data," referred to as *metadata*, can be processed by agents and other application systems to gain access to the *meaning* (semantics) of web content rather than just its *form* (Antoniou & van Harmelen 2008: 10). In web-based learning environments, metadata can be added as semantic markup to web pages and their components in order to create external knowledge representations in different parts of the environment that pedagogical agents can read, manipulate, and reason with. This idea is outlined below and applied in Chapter 16.3.4.

The general format for representing metadata statements is as *triples* consisting of a *subject* (the resource that is the topic of the statement), a *predicate* (what the statement asserts about the subject), and an *object* or *value* (the resource or literal that is the value of the assertion) (Antoniou & van Harmelen 2008: 68ff.). A resource associated with one of these components is identified by a unique name or address called a *uniform resource identifier (URI)* (Berners-Lee et al. 2005), such as `http://www.linguistics-online.com/people/instructors.owl#PF`. For example, the following metadata triple expresses the proposition that a certain individual referred to as "Peter Franke (PF)" (subject) has the current position (predicate) of research assistant (object):

Subject: `http://www.linguistics-online.com/people/instructors.owl#PF`
Predicate: `http://www.linguistics-online.com/people/staff.owl#currentPosition`
Object: `http://www.linguistics-online.com/people/staff.owl#researchAssistant`

The terms used in metadata statements to refer to concepts and relationships of the domain are defined in *ontologies* (Gruber 1993; Uschold & Gruninger 1996; Studer et al. 1998; Vossen 2003), i.e. abstract models which explicitly and formally capture the structure of knowledge in distinct and specified areas, as it is understood and shared by a particular group of people and/or machines. Ontologies usually provide *taxonomies* of domain concepts and rules for deriving new facts through inference. A taxonomy defines a collection of *terms* and their *relationships* that are relevant for describing the domain (Antoniou & van Harmelen 2008: 11). Each term labels a *concept* (*class* of entities) of the domain. Members of a class are called *instances* of that class. The classes are typically organized in terms of *hierarchical* relationships between classes and their subordinate and superordinate classes, which represent more specific and more generic domain concepts, respectively (pp. 85f.). For example, in Figure 59, the class `ACADEMIC STAFF` has three subclasses (`REGULAR FACULTY STAFF`, `RESEARCH STAFF`, and

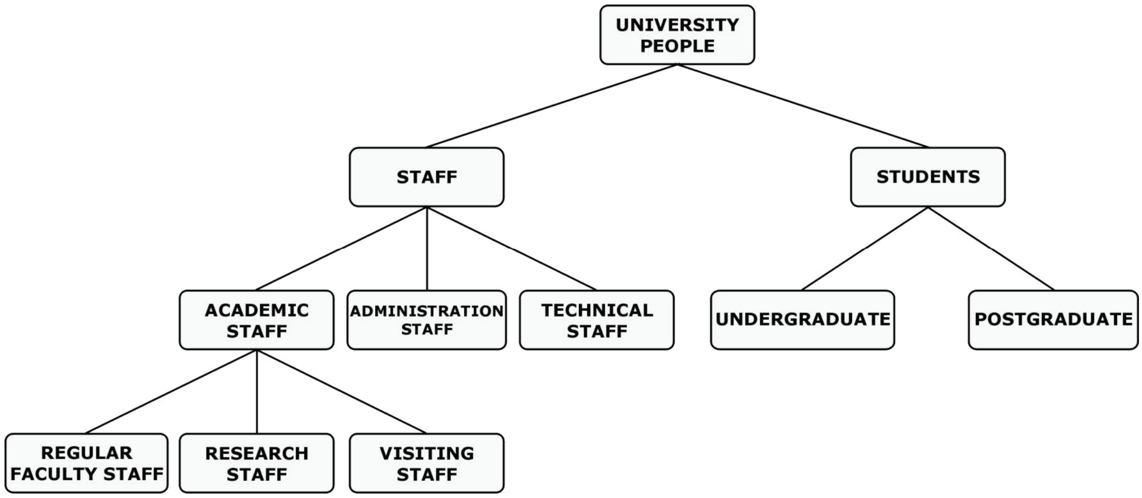

Figure 59. A taxonomic hierarchy representing people at a university. Adapted from Antoniou and van Harmelen (2008: 11, Figure 1.1).

`VISITING STAFF`) and one superclass (`STAFF`). Ontologies may also specify other types of information about their elements, such as *properties* (attributes) of instances of classes, *restrictions* on the range of permissible property values, *disjointness statements* (two classes do not have any common members), and *logical relationships* (synonymy, meronymy, symmetry, transitivity, cardinality, etc.) between ontology elements (pp. 11f.).

One important goal for the Semantic Web is to support the operation of *intelligent software agents* that autonomously and proactively carry out tasks on the user's behalf (cf. Chapter 3), which involves collecting information from various sources on the World Wide Web and elsewhere, communicating with agents representing other individuals or organizations, filtering information according to the user's requirements and preferences, and presenting the user with the result, which may enable him or her to make a choice (Antoniou & van Harmelen 2008: 15). Metadata annotations help agents to find relevant information on the Web and to extract information from web-based resources. Ontologies provide shared vocabularies that facilitate the search for and interpretation of information, as well as the communication with other agents. Logic (cf. Section 12.1.2.1) plays a central role in processing the collected information (including the drawing of inferences) (pp. 15f.). In addition, intelligent agents on the Semantic Web require languages for exchanging messages with other agents (cf. Chapter 14.1.5); formal representations of agents' beliefs, desires, and intentions (including their own) (cf. Chapter 14.1.1); and user modeling capabilities (cf. Section 12.2.1) (p. 16).

For the Semantic Web to emerge, elaborate standards which arise from a consensus among key players are essential (Shadbolt et al. 2006: 99). The architecture of the Semantic Web is being developed step by step in a bottom-up fashion by the various working groups of the World Wide Web Consortium. Each step adds a new *layer* that provides higher-level functionality (Antoniou & van Harmelen 2008: 17–20).[181] The layers of the Semantic Web architecture are shown in Figure 60 (for a detailed discussion of the individual layers, see, for example, Antoniou & van Harmelen 2008). The basic standards of the Semantic Web (URI,

---

[181] It should be noted that the layered architecture shown in Figure 60 is currently the subject of debate in the Semantic Web community. It is quite likely that this architecture will change in the near future to accommodate recent developments (Antoniou & van Harmelen 2008: 20).

XML,[182] RDF,[183] and OWL[184]) are now in place ([INT 127]). Ongoing efforts target the higher levels of rules,[185] logic, and proof (Shadbolt et al. 2006: 98; Antoniou et al. 2007: 62). Various initiatives are creating ontologies for biology, e-government, medicine, environmental science, genomics, and other domains (Neumann et al. 2004; Shadbolt et al. 2006: 96f.). Development tools (for editing, storage, querying and inferencing, visualization, and versioning) have been contributed mostly by academic researchers; however, commercial entities have started to adopt these tools and to contribute their own (Antoniou & van Harmelen 2004: 225). Semantic Web applications of different kinds are being built by both communities.

As indicated at the beginning of this section, in the context of the present work, the Semantic Web approach of annotating resources with machine-processable semantic markup according to a shared conceptualization of the domain (ontology) (Studer et al. 1998) is mainly of interest because it provides an attractive alternative method to represent knowledge for agents operating in web-based environments that involve a limited number of element types whose instances are added, modified, or removed in controlled ways. The elements of such environments (content items, navigation controls, tools, people, agents, etc.) can be conceived of as resources and become the subjects or objects of metadata statements. These statements can be created during the processes of building and maintaining the structure, content, and artifacts of the environment and enrolling its users, ideally with minimal intervention of knowledge engineers (cf. Section 12.1.1.2), content developers, administrative staff, or users because mechanisms embedded in the development tools or in the environment add or change the relevant statements automatically. New objects introduced into the environment may have their own annotations, which integrate seamlessly with the other metadata since they are based on the same domain ontology or on one whose terms can be mapped to it. Ontologies are associated with environments rather than agents, and they are processed by agents to access the vocabulary for

---

182 *XML (eXtensible Markup Language)* is a standardized metalanguage for developing custom markup languages with user-defined tags (e.g. `<firstName>`, `<instructor>`, `<orderNo>`, etc.), which can be used to write structured documents that describe web *content* (rather than layout) (Antoniou & van Harmelen 2008: Chapter 2; [INT 128]). XML provides the syntactic framework for higher-level Semantic Web languages.

183 *RDF (Resource Description Framework)* provides a standardized data model for expressing metadata in a machine-accessible format (Antoniou & van Harmelen 2008: Chapter 3; [INT 129]). Each metadata statement consists of a three-element structure which asserts that the resource *X* has the property *Y* with the value *Z*. RDF statements are commonly written using an XML-based syntax (RDF/XML) (Antoniou & van Harmelen 2008: 66).

184 *OWL (Web Ontology Language)* is the standard language for creating ontologies on the Semantic Web (Antoniou & van Harmelen 2008: Chapter 4; [INT 130]). It provides a layer on top of RDF and RDFS (= RDF Schema, a primitive RDF-based ontology language in which class and property hierarchies can be expressed ([INT 131])) with advanced features allowing the definition of various relationships, restrictions, combinations, and special characteristics of classes and properties. OWL ontologies are written in RDF/XML (Antoniou & van Harmelen 2008: 119). OWL consists of three sublanguages (listed in ascending order of expressiveness and computational complexity): OWL Lite, OWL DL (= "Description Logics"), and OWL Full (pp. 117ff.).

185 *Rules* can be used for various purposes, including knowledge representation, inference, specification of constraints, mapping from stimuli (events, changes, etc.) to responses (actions), and so on. A rule specifies an implication between an antecedent consisting of a conjunction of premises (conditions), and a consequent (Antoniou & van Harmelen 2008: 162f., cf. Section 12.1.2.4). Ongoing work aims to combine OWL and the *Rule Markup Language (RuleML)* ([INT 132]) in order to create a *Semantic Web Rules Language (SWRL)* (Parsia et al. 2005; [INT 133]; Antoniou & van Harmelen 2008: 170f.).

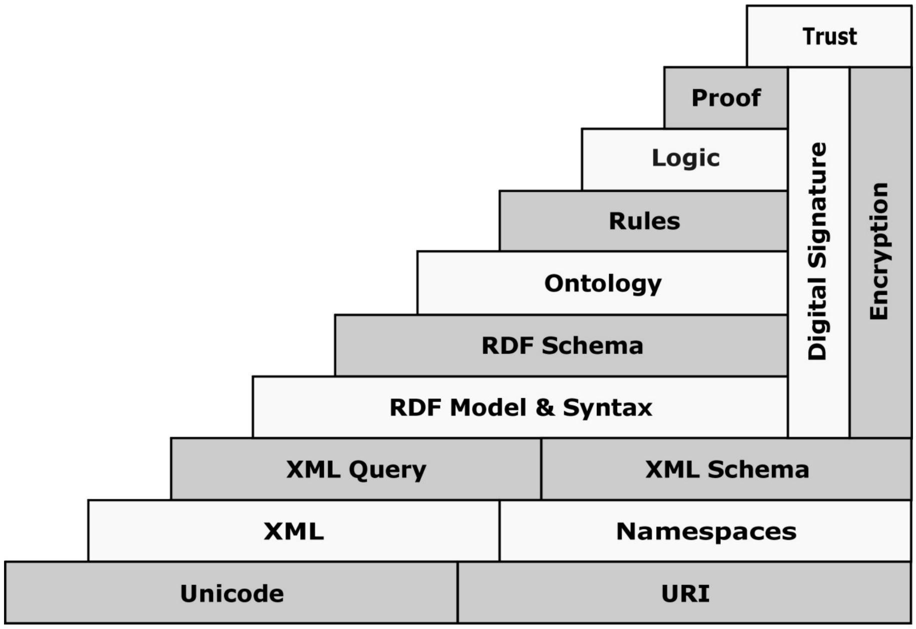

Figure 60. The layered architecture of the Semantic Web. Adapted from ([INT 134]) and ([INT 135]).

describing the domain/environment. Changes to an object can be reflected in its semantic markup, and the object may even carry a history of its changes. Agents interacting with these objects do not have to represent and keep track of knowledge about their properties or status; they only need to be able to read and process the objects' semantic markup. Since the environment contains elements of a limited number of types, the corresponding number of concepts and their relationships in the ontology for the domain is manageable. In sum, at least in theory, Semantic Web technologies can be used to create annotated web-based environments that represent domain knowledge and vocabulary within the environment and in a format that is accessible to agents with different roles, implementations, and origins. In reality, however, these ideas have not yet been widely realized.

The proponents of the Semantic Web regard it as the next important stage in the evolution of the World Wide Web, and they have commonly expected the community of web users to embrace their vision as well and to start contributing, sharing, and reusing metadata and ontologies in order for the Semantic Web to gradually take shape. However, years after the publication of the Semantic Web's inaugural paper (Berners-Lee et al. 2001), none of these things have happened on a large scale yet (Shadbolt et al. 2006; Borland 2007). Instead, these days, Web users collaboratively annotate web-based resources with self-made tags and thus build folksonomies together, both in the context of Web 2.0 (a collective term for recent social and technological developments that are changing the World Wide Web into a medium inviting (and relying on) user participation, user collaboration, and social networking, cf. Chapter 2.3.9). *Folksonomies* can be defined as user-generated taxonomies (classification systems) that emerge organically through a collaborative effort to categorize web resources which involves many individuals who associate these resources with descriptive keywords (*tags*) of their own choice (Shadbolt et al. 2006: 100; [INT 136]; Hayman 2007). While having users create their own descriptive tags and label resources with these tags provides a simple way to define the vocabulary for describing a domain and to generate descriptions of resources, the resulting tag structures are generally not suitable for interpretation by machines, except using statistical techniques like clustering, because formal and unambiguous definitions of the meanings and interrelations of the tags are not available (Shadbolt et al. 2006: 100). As a result, folksonomies may contain ambiguous tags and lack a clear structure (Guy & Tonkin 2006). In short, folksonomies are created for humans, not for machines.

However, it has been suggested that Web 2.0 and the Semantic Web are not incompatible but can complement each other in a hybrid “Web 3.0” (Markoff 2006), where semantic descriptions of web content and services are developed by inspiring and facilitating Web 2.0-typical massive participation of non-expert users, and applications combine Web 2.0 features with RDF and other low-level Semantic Web technologies (Borland 2007; Lassila & Hendler 2007; Ankolekar et al. 2008; Hendler 2008; Wahlster 2008). In these approaches, the more ambitious work on ontologies and inference in the Semantic Web community is largely dismissed as ‘academic’ and thus irrelevant to solving real-world problems (Borland 2007).

In general, the Semantic Web will not achieve its (possibly too) ambitious goal of building “a brain of and for humankind” (Fensel & Musen 2001: 25) by weaving a web of metadata and ontologies, unless there is a critical mass of developers, content authors, web masters, and users that are willing and able to generate and consume these types of machine-readable information. Therefore, the major challenge remains to motivate individuals and organizations to build ontologies and to add semantic markup to the electronic documents, multimedia elements, data, products, and services that they create, host, sell, use, or view on the World Wide Web (or elsewhere), by advertizing and demonstrating the financial, personal, or other gains that can be obtained from the investment of additional resources in the creation of semantic webs (Golbeck et al. 2005: 177). Those motivated to contribute require effective (semi-) automatic tools for producing metadata and ontologies, as well as for linking to existing resources on the Semantic Web. As much semantic markup as possible should be generated automatically, either directly from the source or as a by-product of the user’s regular work (Hendler 2001: 31; Antoniou & van Harmelen 2008: 248). In fact, large-scale automatic annotation of web pages with (simple) semantic markup by means of natural language understanding (cf. Chapter 11.1.3) and machine learning (cf. Section 12.1.1.2) techniques has become feasible (Antoniou & van Harmelen 2008: 248).

Furthermore, the Semantic Web enterprise has to win the *trust* of users, which requires social and technological mechanisms that ensure the validity, identity, and credibility of statements and sources. Logic (cf. Section 12.1.2.1) can be used to automatically prove the *validity* of statements (Hendler 2001: 34f.). *Digital signatures* help to verify the *identity* of sources and to ascertain the integrity of content ([INT 137]). The *credibility* issue is addressed by building a *Web of Trust*, a social network that links each user to a small number of other users he or she trusts (to different degrees). Trust (or distrust) is propagated in a transitive fashion (if *A* trusts *B* and *B* trusts *C* and *D*, then *A* may extend his or her trust to *C* and *D*) through this network (Golbeck et al. 2003; Richardson et al. 2003; Welsh 2003; Golbeck & Hendler 2004; Antoniou & van Harmelen 2008: 19).

Related to trust are concerns about censorship of content and threats to users’ privacy on the Semantic Web (Kim et al. 2002; [INT 138]). Semantic markup is generally invisible (and incomprehensible) to the average user of the Semantic Web. Software (agents) could use this markup to manipulate and pre-select content, to collect sensitive personal information, or to decide what user-related details to share with others, all without the user’s consent or even knowledge. Therefore, the Semantic Web requires effective mechanisms that prevent illegal or unethical manipulation of content and handling of personal data by machines.

12.1.3 Problem Solving

In intelligent agents, knowledge representation is usually not an end in itself but rather plays an important role in enabling these agents to solve problems. But knowledge alone is not enough: intelligent agents also require effective strategies to find solutions using the represented knowledge. For intelligent agents, *problem solving* generally involves *search* (Cawsey 1998: Chapter 4; Russell & Norvig 2003: Part II). That is, the agent has an inventory of actions that *could* be carried out to solve a given problem, but it has no knowledge about which of these actions (in what order) will produce a solution. To find a sequence of actions that solves the problem and meets specified success criteria, the agent has to search through the space of possible action sequences and evaluate different solution candidates (Cawsey 1998: 68; Russell & Norvig 2003: 60). A particular sequence of actions is regarded as a solution if it connects a specified initial state and a desired target state through a series of intermediate states (Cawsey 1998: 69). Problem-solving processes in intelligent agents are subject to various constraints (Russell & Norvig 2003: 35):

- The *goal*, i.e. the target state(s) of the agent and/or the environment to be achieved;
- The *current state* of the agent and the environment;
- The agent's *prior knowledge* of its domain and its environment;
- The agent's *inventory of actions* (i.e. what it can do to affect its environment);
- The agent's *percept sequence* (i.e. the history of its sensor inputs) up to the present moment;
- The *performance measure* that is used to determine the agent's degree of success;
- Limitations of *time* and *resources*.

Given these constraints, a problem-solving agent has to go through the following "formulate-search-execute" cycle (Russell & Norvig 2003: 60f.):

1. *Goal formulation* identifies an objective (desired target state) that the agent will attempt to accomplish.
2. *Problem formulation* determines the set of actions and states that the agent will take into account in trying to achieve the goal.
3. *Search* looks for a sequence of actions that constitutes a solution to the problem (achieves the goal) and satisfies the agent's success criterion, ideally maximizing the performance measure (p. 36).
4. *Execution* performs the action sequence specified by the solution found.

12.1.3.1 Search Techniques

When computers are faced with the task of finding a solution to a problem, one typical strategy involves generating all possible solutions (i.e. sequences of actions) and then searching through the whole set or a subset of these possibilities to identify one (or several) that meet specified requirements. The set of solution candidates to be considered may be large, so efficient *search techniques* are needed.

If the problem is of low complexity, a systematic and exhaustive search of the entire set of possible paths is feasible. The search process examines every state that can be reached from the initial state (on a path leading through one or several intermediate states). If it is a target state, a solution has been found. The set of all states to be considered in a given search problem is called the *search space* (Cawsey 1998: 69). In the following, the simplifying assumption is made that the search space is structured as a search tree, in which each node (state) can be reached from the root node by only one path (p. 70).[186] Search techniques that systematically consider all the paths in the search space are called *brute force* or *blind* search techniques (pp. 69f.). The simplest two among the blind search techniques are:[187]

- *Breadth-first search*. Explore all states in a search tree at a given depth before moving on to their direct successors at the next level. In other words, the tree is searched one level after the other (Cawsey 1998: 72ff.; Russell & Norvig 2003: 73f.).
- *Depth-first search*. Explore a given branch of a search tree to its full length before exploring other branches (Cawsey 1998: 74f.; Russell & Norvig 2003: 75ff.).

Brute force techniques may be too inefficient for complex problems that involve large search spaces. For such problems, *heuristics* (rules of thumb that allow an informed guess of what the next problem-solving step will be) can indicate which paths most likely lead to a solution (Luger 2005: 123f.). Rather than search through the whole space of possibilities, only those paths or states that appear promising according to a heuristic *evaluation function* $f(n)$ are considered. This function estimates the distance from any state n of the search space to the target state. Those states (or paths) that move the problem-solving process closer to the goal state are tried before considering the others (Cawsey 1998: 77). For example, *best-first search* uses an evaluation function to determine the state with the best score (i.e. smallest distance to the goal) and always selects that state as the next one (pp. 79f.). The most widely used best-first search algorithm is *A* search*, which attempts to find a solution path through state n that minimizes total cost as defined by the evaluation function $f(n) = g(n) + h(n)$, where $g(n)$ is the actual cost (distance) of reaching n from the initial state and $h(n)$ is the estimated cost of getting from n to the target state. The state n with the lowest score for $f(n)$ (and thus the lowest estimated total cost of the solution path through n) becomes the next state (Cawsey 1998: 80ff.; Russell & Norvig 2003: 97–104). Provided that a good evaluation function is used, the score produced by that function, while only a guess, can help to speed up the search process considerably (Cawsey 1998: 77 + 80). However, heuristic search algorithms (Cawsey 1998: 78–82; Russell & Norvig 2003: Chapter 4; Luger 2005: Chapter 4) do not reliably produce the best solution to a problem. In fact, they will quite often find suboptimal solutions or even none at all (Luger 2005: 124).

Search algorithms are basic components of various advanced problem-solving techniques. Three of them, state space search, planning, and practical reasoning with beliefs, desires, and intentions, are discussed in the following.

12.1.3.2 State Space Search

State space search (Cawsey 1998: 82–87) is a widely used search-based method for problem solving. Problems are specified in terms of an initial state and a target (or goal) state. Solutions are action sequences that link these two states (via any number of intermediate states). The

[186] In general, search graphs may have a more complex structure with multiple paths leading to a given state. This raises the problem of how to avoid considering the same state repeatedly. The detection of repeated states requires the search algorithm to maintain a history of previously visited states. If the current state is found in this history, it is discarded (Russell & Norvig 2003: 81ff.).

[187] Further, more sophisticated blind search algorithms based on breadth-first or depth-first search are discussed by Russell and Norvig (2003: 73–81).

nodes in the search space correspond to different states of the problem (Cawsey 1998: 82). The search space is also called the *state space* because it consists of all states that can be reached from the initial state (through other states) (Russell & Norvig 2003: 62). A limited set of (abstract or physical) actions is available to solve the problem. Each action is represented as a rule or operator that, if applied, changes the problem state in a well-defined way (Cawsey 1998: 82f.). Starting from some initial state, different paths are constructed by identifying actions applicable in the current state and using them to determine the states that result from them (in other words, possible extensions of the path that led from the initial state to the current state). By systematically searching through all possible paths (using blind or heuristic techniques) and discarding those that contain loops, a path may be found eventually that connects the initial state and the target state (pp. 83–85).

12.1.3.3 Planning

Another approach to problem solving constructs *plans*, i.e. (incomplete) specifications of future courses of action (that of the reasoning agent or some other individual), which determine the agent's own behavior and help the agent to interpret or adapt to the behavior of other entities (Tate 2001: 653). Plans break some high-level action down into a hierarchy of successively more specific subactions. Each level of the hierarchy shows the temporal order in which the actions specified at that level have to be carried out in order to achieve goals and subgoals (cf. Figure 52 and Chapter 11.3.3.2.2). The *planning* process (Cawsey 1998: 87–90; Russell & Norvig 2003: Part IV) applies search techniques to find a decomposition of higher-level actions into lower-level actions which at the lowest level consists of a sequence of primitive actions that will achieve the specified goal. This sequence represents a possible plan of action, which may be evaluated and finally executed (Cawsey 1998: 87; Russell & Norvig 2003: 375). Planning typically involves building alternative plans and reasoning about their consequences.

One early, simple planning technique is based on the *means-ends analysis (MEA)* approach (Copeland 1993: 86–91; Cawsey 1998: 87ff.). Means-ends analysis is a control strategy that regulates the search process involved in finding a path from an initial state to a desired goal state (Cawsey 1998: 87; [INT 139]). It examines the actions (means) available to accomplish goals (ends) and selects those actions that, given the current state, result in a new state which differs less from the goal state. If such an action is not applicable in the current state, the planner is given the new subproblem of reaching a state that allows the application of this action (Cawsey 1998: 87). The actions can be modeled as plan operators specifying the pre-conditions of an action and its effects on the problem state (p. 88). A given plan operator is considered for application if its effects bring about a new state that is more similar to the target state. The pre-conditions of the operator determine its applicability in the current state. If certain pre-conditions are not fulfilled, they become goals with which the means-ends algorithm is called recursively (p. 88f.).

Means-ends analysis has provided the basis for several more advanced planning approaches, which include:

- *Planning with goal protection*. Previously achieved goals are preserved by ensuring that they are not accidentally undone by the action currently under consideration (Cawsey 1998: 90).
- *Partial-order planning*. Instead of requiring all actions in a plan to occur strictly one after the other, it is possible to include actions whose temporal order is not specified, provided that one action does not depend on the other (Cawsey 1998: 90; Russell & Norvig 2003: 388).

- *Hierarchical planning*. Planning starts from a high-level initial plan that is successively refined through decomposition of higher-level actions into lower-level and, eventually, primitive actions (Russell & Norvig 2003: 422f.).
- *Reactive planning*. Instead of a complete action sequence, the planner only determines the next action to be carried out, given the current context. As a result, it becomes possible to deal with environments that are dynamic and unpredictable ([INT 140]).[188]

12.1.3.4 Practical Reasoning

Practical reasoning is the rational process by which intelligent agents work out, moment by moment, what to do in order to get closer to achieving their goals (Bratman 1987; Wooldridge 1999: 54). This involves determining what goals to achieve (deliberation) and how to achieve them (means-ends reasoning) (Wooldridge 1999: 55). Practical reasoning agents first generate a set of options (i.e. alternative courses of action) available to them, then make a choice between these options, weighing their advantages and disadvantages, and finally commit to the selected options as *intentions* that drive their actions and influence their future practical reasoning (p. 55). Intentions serve several important functions in practical reasoning agents (pp. 55ff.):

- *Intentions drive means-ends reasoning*. Agents will try to achieve an intention to which they have committed themselves, which involves deciding on an appropriate course of action. If the chosen option fails, another may be attempted.
- *Intentions constrain future deliberation*. Options that are not consistent with a given adopted intention will not be considered by the agent.
- *Intentions persist*. They will not be abandoned unless either the agent has achieved them, the agent comes to believe that they are unattainable, or their justification is gone.
- *Intentions influence beliefs upon which future practical reasoning is based*. For example, the agent should believe (with a certain degree of confidence) that it will achieve its intentions. The agent cannot intend to achieve something and at the same time believe that this will never happen. Such behavior would be inconsistent and irrational.
- *Intentions involve a tradeoff between the degree of commitment and reconsideration*. If the environment is static, bold agents (which never reconsider their intentions) perform better than cautious agents (which reconsider their commitments all the time) because they stay on target without being distracted by unnecessary reasoning about whether they have to modify their intentions. By contrast, in dynamic environments, cautious agents are better off because they can recognize unsuccessful intentions and seize opportunities as they arise.

The *belief-desire-intention (BDI) architecture* is possibly the most widely used model of practical reasoning agents. The BDI model will be discussed in Chapter 14.1.1.

12.2 The Student Model

The ability of pedagogical software agents to effectively tailor their instruction and other kinds of assistance to the individual learner depends critically on the image, or model, of that learner which they have been able to construct in the course of a history of learner-agent interactions.

[188] See also the discussion of reactive agent architectures in Chapter 14.1.2.

This section reviews techniques that enable pedagogical agents to become *learner-*, or, more generally, *user-adaptive*. First, the more general task of modeling the user in user-adaptive systems is discussed in Section 12.2.1. The following sections are then concerned with user modeling for educational purposes in an educational context, which is commonly referred to as *student modeling*. Section 12.2.2 provides a discussion of the nature, contents, and functions of student models, as well as of issues involved in their design. Section 12.2.3 reviews different approaches to student modeling that have been described in the literature, chiefly for intelligent tutoring systems (cf. Chapter 2.3.3).

12.2.1 User Modeling

State-of-the-art computer-based applications are commonly characterized by *adaptability*. An adaptable system provides the individual user with the means to adjust its appearance and behavior so that it better suits his or her preferences (Jameson 2003: 306; Rothrock et al. 2002; Johnson & Taatgen 2005: 430). While the possibility to adapt a given system emphasizes users' freedom of choice and promotes their sense of being in control of the system, it also has a number of limitations (Jameson 2003: 308):

- The user may not know what aspects of the system can be adapted and how this can be done.
- The user may be overwhelmed by large numbers of options.
- The user may have to find his or her optimal personal settings by trying many different ones.

A different perspective on adaptation stresses the active role of the machine in the process: rather than have the user adapt the system, the system adapts to the user. This ability is a key feature of *user-adaptive systems (UASs)* (McTear 1993; Kobsa 2001; Rothrock et al. 2002; Jameson 2003; Johnson & Taatgen 2005), which include (pedagogical) software agents (cf. Chapter 3.1.2). A user-adaptive system (or agent) has been defined as:

> *An interactive system that adapts its behavior to individual users on the basis of processes of user model acquisition and application that involve some form of learning, inference, or decision making. (Jameson 2003: 306).*

The general motivation for incorporating models and processes enabling adaptation to users into agents and other interactive systems is to help the individual user to achieve his or her goals through the interaction with the system more effectively (cf. Chapter 6.3) and to make the experience more enjoyable for the user (cf. Chapter 6.4). Jameson identified several functions that can be served by user adaptiveness, grouped into functions supporting the use of the system (doing (parts of) routine tasks for the user, tailoring the user interface, providing adaptive help, and choosing appropriate dialogue strategies (cf. Chapter 11.3.3.2.1)) and functions facilitating individual or collaborative acquisition of information (recommending or selecting information; customizing information presentation; predicting user interests; supporting collaboration among users; and adapting the content, form, selection, and timing of instructional interactions (cf. Section 12.3)) (Jameson 2003: 307–316).

Processing in a user-adaptive agent involves two major steps: the *acquisition* of a *user model* from information about the user and the *application* of the user model to make predictions or decisions about the user (Jameson 2003: 306, cf. Figure 61). In the acquisition step, the agent applies machine learning (cf. Section 12.1.1.2) or inference procedures (cf. Section 12.1.2.1) to information collected about the user in order to build the user model (p. 306). The outcomes of the acquisition process are explicit assumptions about various user-related aspects (Wahlster

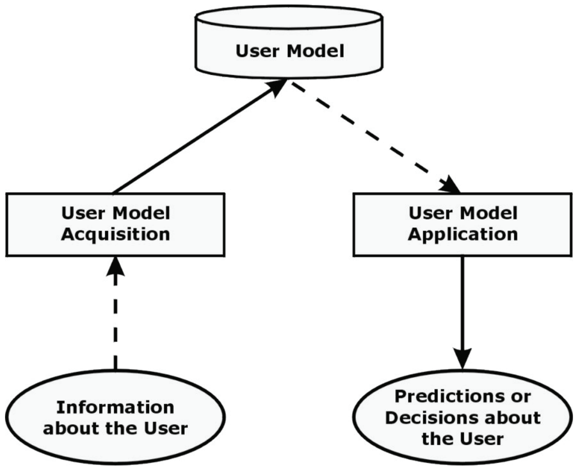

Figure 61. Processing in a user-adaptive system. Adapted from Jameson (2003: 306, Figure 15.2). Dotted arrows indicate that certain information is used by a component process, whereas solid arrows link processes to their outcomes.

1988: 102) that may serve the agent in carrying out its tasks and building a relationship with the user (cf. Chapter 7.16). The assumptions in the user model are applied in the analysis of the current agent-user situation in order to inform (decisions about) adaptive responses of the agent (Jameson 2003: 306).

Various aspects of the user may be of interest to a user modeling agent. The following is an incomplete list of user characteristics (cf. Chapter 6.1) that may be included in the user model of an adaptive agent:

- Goals and plans (for the interaction and beyond);
- Expectations (of interaction and performance, cf. Chapter 6.1);
- Background (cultural, demographic, economic, educational, etc.);
- Expertise level (linguistic and non-linguistic knowledge and skills);
- Preferred mode of interaction (spoken, written, or multimodal, cf. Chapter 11);
- Limitations (originating both from within the user and from the context, cf. Chapter 6.1);
- Tasks (to be accomplished by the user alone or in cooperation with the agent);
- History of interactions (including conversation logs, results, and user feedback);
- Perceived emotional state (e.g. angry, depressed, happy, cf. Chapter 13.1);
- Identified personality traits (e.g. extrovert vs. introvert, cf. Chapter 13.5);
- Degree of closeness (to the agent and other individuals);
- Relative social status (with respect to the agent and other entities).

The process of acquiring a user model may draw on both information explicitly provided by users and information obtained unobtrusively by means of more indirect techniques. The former approach elicits *self-reports* and *self-assessments* from users that inform the adaptation process (Jameson 2003: 318f.). Many user-adaptive systems ask their users to provide their age, gender, profession, and other ‘objective,’ relatively static personal information (p. 318). Furthermore, it is possible to have users rate themselves on various dimensions, such as how much they are interested in or know about a given product, topic, etc. or how important they deem a certain feature, issue, or criterion (p. 318f.). Apart from self-rating, users can also be asked to evaluate specific items or actions (p. 319). In computer-based learning environments, learner assessment activities are common practice and thus provide natural opportunities for collecting information about each learner's progress, misconceptions, gaps in knowledge, etc. (p. 319). While the direct approach is generally easy to implement using forms and questionnaires (Johnson & Taatgen 2005: 430, cf. Chapter 15.4.5.1), it is also obtrusive (distracting the user from what he or she is trying to accomplish) and threatens the user's privacy (stored user-related information may be used for purposes that are not in the user's interest and do not have his or her approval) (Jameson 2003: 317).

Instead of asking (and possibly interrupting) the user, a user-modeling agent equipped with an appropriate array of internal and/or external sensors (cf. Chapter 4.2) can less obtrusively monitor the user and his or her behavior while he or she is interacting with the agent or system, and interpret the collected data in terms of various user-related dimensions (see above). This approach has the advantage that users do not have to interrupt their normal work and perform actions that are only relevant for user modeling; they just do what they always do. On the downside, the data generated by monitoring natural user behavior is often incomplete or ambiguous (Jameson 2003: 320). Furthermore, there is the issue of whether or not to reveal to users that they are being monitored. The most authentic data can be obtained if users are unaware of the agent's monitoring because if they knew, this would quite likely influence their behavior. For example, users might feel uneasy about being watched and behave accordingly (cf. Chapter 10.4), or they might try to present themselves in a certain (positive) light. However, it is also true that users have a right to know what data a monitoring agent is collecting about them and for what purposes, and that users generally do not appreciate it if such things are happening behind their backs (p. 320).

Recently, the possibilities of user monitoring have been significantly extended because of the availability of increasingly small and sophisticated external sensors that can capture low-level physiological signals from users, in particular signals related to the bodily experience of emotions (cf. Chapter 13.1.1). User-modeling agents can now collect quite reliable data about the physiological correlates of anger, frustration, boredom, attraction, stress, and other mental states of the user (Jameson 2003: 320). Physiological data, such as blood pressure, body temperature, and respiration patterns, can be obtained both from sensors that are attached to (or, in the future, may even be implanted in) the user's body and from sensors that are embedded in objects and devices that have physical contact with the user (including clothing) (Picard 2000; Jameson 2003: 320). Cameras and microphones can monitor the user's face (cf. Chapter 10.1.3) and voice (cf. Chapter 10.1.1), respectively, for additional cues as to what he or she is thinking or feeling (Jameson 2003: 320, cf. Chapter 13.4.1). The incoming data from the different sensory channels is analyzed and fused into a meaningful interpretation of the user's current state, possibly using disambiguation information from other sources (p. 320, cf. Chapter 11.2). All the sensors mentioned above can provide the user modeling process with an uninterrupted supply of data whose corresponding bodily responses are not (fully) under the user's conscious control (cf. Chapter 15.4.5.5). As a result, assessments of his or her internal state may be closer to reality. However, again, privacy issues arise: users may feel threatened by machines that can probe deeper than the average person.

Users typically interact with several computer-based systems, and other systems already may have acquired valuable information about the user. Therefore, drawing on stored information previously collected by others can be a feasible strategy for user-modeling agents (Jameson 2003: 320). Examples of *external information sources* include (p. 320):

- Records about the user (as a customer, patient, student, etc.) in databases maintained by the institution or system of which the user-modeling agent is also a part;

- Information in publicly accessible repositories, in particular the World Wide Web;
- Results of user modeling by other agents or systems.

Information from external sources is available to an agent already on its first encounter with the user. However, the content and form of the external information may limit its usefulness for the agent's user modeling efforts. Furthermore, accessing confidential user-related information without the user's knowledge or permission is unlikely to inspire trust or liking when the user finds out, and may also have legal consequences (Jameson 2003: 320).

Finally, user-adaptive agents can extend their attention beyond the individual user to include the *context of the interaction* (Jameson 2003: 320f.). The context includes other individuals (both human and artificial), the task to be accomplished, artifacts relevant to the interaction, and the current physical, sociocultural, and organizational environment (cf. Chapter 6.2).

In general, it could be assumed that more, and more diverse, information about a particular user can only help a user-modeling agent in its attempts to adapt to that user's behavior or characteristics. However, the usefulness of the collected information depends on its relevance, detail, quality, and timely availability, as well as on the agent's ability to make use of the information. In fact, it might happen (and has happened before) that much more user-related information is being collected than the agent can use (cf. Chapter 12.2.2).

User-adaptive agents need computational techniques that help them to learn different things about the user, to draw conclusions about the user from their knowledge, and to decide on the best way to adapt to the user (Jameson 2003: 321). Techniques from different computational paradigms have been used for this purpose (pp. 321–324). One of them, *classification learning*, subsumes various techniques from the field of machine learning (cf. Section 12.1.1). All of these techniques offer solutions to the basic problem of learning a *classifier*, i.e. a model that can assign an unseen item to one of a set of pre-defined categories based on the item's features, from a collection of training data whose items were previously classified into the same set of categories (Cawsey 1998: 179; Jameson 2003: 321). Often, there will be uncertainty about how to correctly classify a given item. In such cases, a set of possible classifications ranked according to confidence may be returned (Jameson 2003: 321).

Webb et al. discussed several issues faced by machine learning in user modeling. First, the machine learning techniques involved require large sets of training data that have been explicitly and accurately labeled. Second, the computational complexity of the trained classifiers should be acceptable for practical systems. Third, the features used to model the user may change with the passage of time (Webb et al. 2001).

Popular examples of classification-based techniques include:

- *Decision trees*. These flow-chart-like tree structures consist of tests or questions (nodes), answers (branches), and conclusions (leaves). A decision tree is traversed by answering question after question until the leaf node with the conclusion is reached (Cawsey 1998: 180).
- *Probabilistic classifiers*. These classifiers compute the *conditional* (or *posterior*) *probability* $P(c_i|d_j)$[189] that a pattern d_j consisting of a set of feature variables $f_{1j},\ldots,f_{nj}$ falls into a particular category c_i, using Bayes' theorem ([INT 141]). Based on the assumption of *conditional independence* between feature variables, the widely used *Naïve Bayesian classifiers* calculate this probability as $P(c_i|d_j) = P(c_i)P(d_j|c_i)$, where $P(c_i)$ is the *unconditional* (or

[189] The conditional or posterior probability $P(A|B)$ of "A given B" is the probability of an event A given knowledge that another event B has occurred (Russell & Norvig 2003: 470f.; [INT 142]).

prior) *probability*[190] that an item picked at random can be classified into category c_i, and $P(d_j|c_i) = P(f_1|c_i)P(f_2|c_i)...P(f_n|c_i)$, i.e. the product of the individual conditional probabilities of the feature variables (Russell & Norvig 2003: 481f.; [INT 141]).

- *Case-based reasoning (CBR)*. In this approach, a new problem is solved by 1. *retrieving* the most similar previous problem situations (*cases*); 2. *reusing* the specific knowledge of these cases in the new problem situation; 3. *revising* the suggested solution; and 4. *retaining* information from the current case that may be of interest in solving future cases (Aamodt & Plaza 1994).
- *Artificial neural networks (ANNs)*. These networks consist of a large number of processing units that are linked by connections whose weights are determined by training. Processing arises from the interaction between the connected units, all of which are capable of operating in parallel (Handke 1995: 44; Holmes & Holmes 2001: 214, cf. Chapter 11.1.1.2).
- *Text classification*. This technique is used in information retrieval to automatically induce a model of the user's interests or information needs (by using a machine learning algorithm to learn a classifier based on a collection of user-rated documents) with the goal of improving document retrieval (Webb et al. 2001: 23).

Collaborative filtering approaches anticipate users' responses to items (e.g. their ratings of products) based on how other users with similar response patterns ('neighbors') have responded to the same items before, rather than on characteristics of the items themselves (Jameson 2003: 322; Johnson & Taatgen 2005: 430; [INT 144]). Large data sets of user responses to relevant items, preferably by users who are similar to the current user, are required to allow reliable predictions ([INT 144]). However, the accumulation of these responses may take more time than is appropriate for applications involving a high fluctuation of items (Jameson 2003: 322). Users may also have to contribute a substantial number of responses before the system is able to find similar users, predict the user's preferences, and make useful recommendations (Jameson 2003: 312 + 322; [INT 144]).

The third paradigm, *decision theory*, is concerned with goal-directed behavior (choices) of rational agents (including both humans and intelligent machines) given closed sets of mutually exclusive alternatives, i.e. it deals with how agents should make decisions to qualify as rational (*normative*) or how agents actually make decisions (*descriptive*) (Hansson 2005: 6). Decision making processes are influenced by how much an agent knows about the probabilities of the outcomes of the different alternatives. The spectrum ranges from *certainty* (known outcomes) to *risk* (known probabilities) to *uncertainty* (partial knowledge of probabilities) to *ignorance* (no knowledge of probabilities) (pp. 27f.). A rational agent facing risk in a decision situation calculates the *expected utility*[191] of each alternative as the weighted sum of the utility values associated with the different possible outcome states of the alternative, using the probabilities of these outcomes as weights on the utility values, and picks the alternative with the highest (maximum) expected utility (Hansson 2005: 29; Russell & Norvig 2003: 585). Typically, the probabilities represent an individual's degree of *belief* in a state (or event), given his or her background information (Horvitz et al. 1988; Russell & Norvig 2003: 464; Hansson 2005: 37). Likewise, the utility of different options usually also depends on the decision maker's point of view (Hansson 2005: 30f.). The calculation of expected utility based on both subjective utilities

---

[190] The prior or unconditional probability $P(A)$ of a random event A is its probability if no other information is available (Russell & Norvig 2003: 468ff.; [INT 143]).

[191] The expected utility of a given alternative is a numerical value indicating the degree of happiness or gratification that a decision-making agent can obtain when selecting that option.

and subjective probabilities is at the heart of *Bayesian decision theory* (p. 37), which models decision situations in terms of a set of *alternative actions* (physical actions, communicative acts, action sequences, etc.), a *probability distribution* assigning a probability to each possible world state, the *outcome* of each action in each possible state, and a *utility function* quantifying the degree of usefulness of each possible outcome (Haddawy 1999). A Bayesian decision-theoretic agent decides on an action, given a set of probabilistic beliefs about the present world state, by (1) calculating updated probabilities for the current belief state; (2) computing outcome probabilities for actions; and (3) selecting the action (or one of several equivalent actions) that *maximizes expected utility* (Russell & Norvig 2003: 466).

The central tool for representing and reasoning about decision problems in the Bayesian framework is a graphical network model that describes probabilistic relationships among variables[192] (Heckerman 1996). *Bayesian networks* (or *belief networks*) (Pearl 1988; Charniak 1991; Heckerman 1996; Jensen 1996; Haddawy 1999; D'Ambrosio 1999; Pearl 2001; Pearl & Russell 2003; Russell & Norvig 2003: Chapter 14; Ben-Gal 2007; [INT 145]), such as the partial network in Figure 62, are *directed acyclic graphs (DAGs)* composed of nodes that correspond to random variables of interest (in user-adaptive systems typically (observable or unobservable) properties of the user) about which an individual or system has a probabilistic belief, and directed arcs (or links) that encode causal or informational dependencies between the connected nodes/variables (Pearl 2001; Russell & Norvig 2003: 493; [INT 145]). Each node X_i with inbound links is associated with a conditional probability distribution $P(X_i|Parents(X_i))$ that gives the probability of each possible value of the node for each possible combination of the values of its *parents* (i.e. nodes that have an outgoing directed link to the current node (one of their *children*)) (Haddawy 1999; Murray et al. 2001; Zukerman & Albrecht 2001: 11; Russell & Norvig 2003: 493). Nodes without parents are associated with prior probability distributions (Murray et al. 2001). The prior and conditional probability distributions are also called the *parameters* of the network (Michel & Jacobs 2007: 12).

The interlinked organization of Bayesian networks makes it possible to compute (updated) probabilities for any subset of variables in a network based on information available about any other subset (Pearl 2001). If the value of a particular variable is observed, the corresponding node in the network is *instantiated* (i.e. assigned the observed value with certainty), and the evaluation of the network leads to revised beliefs about the other variables (Mayo & Mitrovic 2001: 126). The possibility to spread information through a network in any direction (Pearl & Russell 2003) enables different types of inference. For applications in user modeling, it is important that Bayesian networks support *predictive inference* (from causes to possible effects or actions) through downward propagation of beliefs from parent to child nodes in the network, as well as *diagnostic inference* (from observed effects or actions to possible causes) by way of upward propagation in the reverse direction (Jameson 1995: 196f.; Haddawy 1999; Lauría & Duchessi 2006: 1575). In a *causal Bayesian network*, the parents of a given node are interpreted as its immediate causes (Pearl & Russell 2003). For example, the network in Figure 62 shows that the user's need for assistance, on the one hand, depends on his or her expertise and the difficulty of the current task, and, on the other hand, accounts for certain user behaviors (menu surfing and pauses). Furthermore, task difficulty is represented as a possible cause of user distraction which, in turn, is a further likely reason for the user pausing after some activity (Jameson 2003: 323).

[192] Examples of variables include the temperature of an engine, the test score of a student, the rating of a product, the occurrence of an event (Pearl 2001), the user's need for assistance (Horvitz et al. 1998), etc.

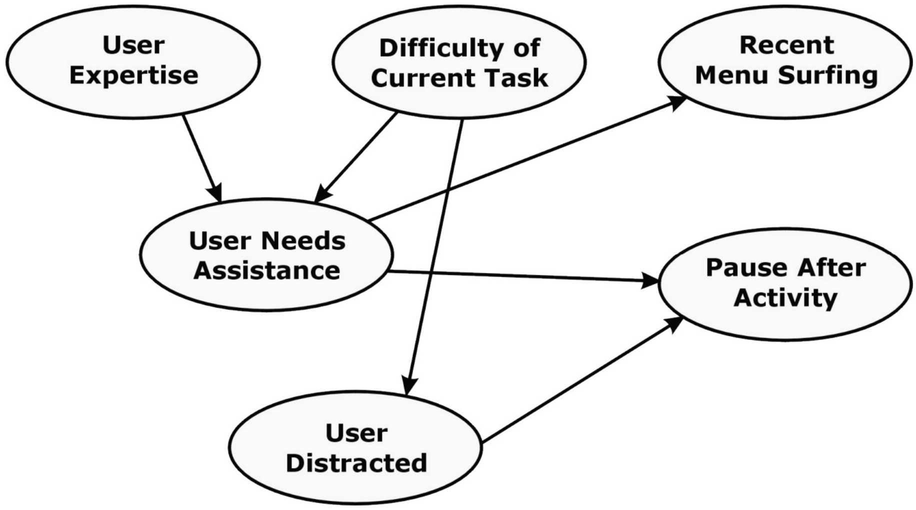

Figure 62. A fragment of a Bayesian user model (adapted from Horvitz et al. (1998, Figure 2)) that allows to infer the user's need of assistance based on his or her expertise and the difficulty of the current task as well as on observations of the user's recent activities (menu surfing and pauses) (Horvitz et al. 1998).

Jameson argued that the use of Bayesian networks and other decision-theoretic models in user modeling has several advantages. For one, the modeling agent does not have to engage in long-term collection of data about a particular user before it can draw conclusions about him or her. Furthermore, decision-theoretic models can help agents to select adaptive actions by weighing the quantified anticipated benefits and costs of different options (Jameson 2003: 323). However, according to Jameson, the difficulty lies in the construction of decision-theoretic models, which become complex for realistic applications and require the manual specification of their parameters and structure (p. 323). But there are now several methods for learning these aspects of decision-theoretic models from data (e.g. Heckerman 1996; Russell & Norvig 2003: 712–733). For further discussion of Bayesian networks, see Section 12.2.3.4 on probabilistic student modeling.

In the *plan recognition* paradigm, the observed actions of users are used to infer their goals as well as the plans by which they intend to achieve these goals (Carberry 2001: 31; Jameson 2003: 323). Once a user-adaptive agent has uncovered the user's plan, it can use this knowledge to help the user to carry out his or her plan, by asking if it should take over the execution of routine plans in the future, by diagnosing the user's plan and offering helpful suggestions, or by planning its next dialogue move or other action so as to further the user's plan (cf. Chapter 11.3.3.2.2) (Jameson 2003: 323).

Depending on the role played by the observed agent in the recognition process (here: the user), two major types of plan recognition can be identified (Carberry 2001: 32): *keyhole recognition*, in which the agent (user) does not know (or does not care) that he or she is being observed (in other words, the observation is unobtrusive, and the user does not try to influence the process of recognizing his or her plan), and *intended recognition*, in which the observed agent is assumed to attempt to facilitate (or hinder) the recognition of his or her plan by deliberately structuring his or her activities in a certain way (Cohen et al. 1982). Given a set of potential goals of a user in the domain and an observed action of the user, plan recognition for user modeling must determine the user's goal and how it is furthered by the observed action (Carberry 2001: 32). The plan inference system knows what user actions are possible in the domain and has a *plan library* which represents each action in terms of a *plan operator* specifying the action's preconditions and the effects (or goals) of its execution (p. 32, cf. the examples in Chapter 11.3.3.2.2). The user's goal is inferred from his or her observed action by constructing a chain from actions to their goals (effects) to other actions for which the goals are preconditions, and so on, starting from the observed action and ending in a possible domain goal. The resulting inference chain of goals and actions accounts for the user's observed action by making its contribution to the domain goal explicit (p. 33). In the real world, multiple chains (hypotheses about the user's plan) with different degrees of plausibility may be constructed for a given observed user action (p. 34). To eliminate less plausible hypotheses, the plan inference system could ask the user for clarification, apply heuristics based on rationality or coherency, use domain knowledge (cf. Section 12.1), or exploit knowledge about the user or group of users (pp. 34ff.).

The final user-modeling paradigm to be discussed here assigns a given user to one or several of a set of pre-defined user categories (*stereotypes*), using a first impression of the user based on, for example, observations of user behavior, users' self-reports and self-assessments (see above), or knowledge about what kinds of people typically use the system (Jameson 2003: 324; Tsiriga & Virvou 2003). Each stereotype provides a set of default assumptions about the user that can provide the basis for constructing a more elaborate user model (Kay 2000). *Triggers* associated with a stereotype specify a set of conditions under which to activate the stereotype for the current user. The *body* of the stereotype provides defaults that are assumed to apply to users who fit the stereotype (Rich 1989). Acquiring an (initial) model for the current user then involves activating all stereotypes with satisfied trigger conditions and applying the information in their bodies to the user. Stereotypes are an old but still very popular user modeling technique (Jameson 2003: 324; Tsiriga & Virvou 2003). Only a small amount of information about the user has to be collected in order to trigger a stereotype (Kay 2000), and the default assumptions provided by the stereotype can either directly serve as the user model or provide the basis for further modeling. Of course, as noted in Chapter 7.15, stereotypes are simplifications and thus may not do a particular user justice. Hence, a stereotype-based user model may be too coarse-grained and require further elaboration.

12.2.2 Aspects of Student Models

Student modeling is the continuing process of building and maintaining "an approximate, possibly partial, primarily qualitative representation" (Sison & Shimura 1998: 131f.) of an individual learner's goals, plans, knowledge, beliefs, attitudes, emotions, personality, etc. (the *student model*), to the extent to which information about these aspects can be obtained with reasonable computational effort by diagnosing and making inferences from the learner's observable behavior (pp. 129 + 131f.). Much of the work on student modeling so far has been done in the field of *intelligent tutoring systems (ITSs)* (Sleeman & Brown 1982; Wenger 1987; Self 1988; VanLehn 1988; Rickel 1989; Nwana 1991; McCalla 1992; Brusilovsky 1994; Holt et al. 1994; Self 1994a; Shute & Psotka 1996; Self 1999), which has used artificial intelligence techniques to build computer-based tutors that are similar to good personal human tutors (Servan-Schreiber 1987) in their "ability to perceive the student's view and to adapt [their] behavior accordingly" (Wenger 1987: 16, cf. Chapter 2.3.3). The case for providing individualized learner support is usually made by pointing out that students being tutored individually have been shown empirically to outperform those being taught in a group setting by up to two standard deviations (the so-called *two-sigma effect*) (Bloom 1984). ITS developers have tried to achieve this effect with their systems but have not yet succeeded (Abdullah 2003: 12).

Student models are primarily *qualitative* in nature, i.e. they use spatial, temporal, or causal relations, rather than numbers, as their major descriptive devices (Clancey 1986: 394; Sison & Shimura 1998: 132), although they may contain quantitative information like scores and frequencies. They are typically *approximate* and *partial* because a complete and precise model of all aspects of individual students' behavior and characteristics is neither computationally (or theoretically) feasible nor necessary in practice (Sison & Shimura 1998: 132). In other words, computational utility of student models is more desirable than their cognitive fidelity (Self 1994a).

In contrast to other components of pedagogical agents, the student model is not a static pre-built resource. The modeling agent dynamically constructs its model of the individual learner during its interactions with the student, using the accumulating evidence obtained by recording, diagnosing, and concluding from the learner's activities, expressions, responses, and other cues. The current student model allows the agent to adapt its instructional interactions (cf. Section 12.3) to each learner, which includes:

- The design, delivery, and timing of interventions;
- The generation of individualized feedback and problem-solving hints;
- The identification and remediation of errors, gaps, and inconsistencies in learners' knowledge and skills;
- The detection of opportunities for teaching the learner new knowledge and skills;
- The communication of encouragement, empathy, praise, etc.

In short, paraphrasing Self, a student model is what enables a pedagogical agent to *care* about each individual learner (Self 1999: 352). Four central questions have to be answered in building such a model (Stauffer 1996):

1. *Who* will be modeled? – The degree of specialization with respect to the object of modeling (a typical learner, different groups of learners, or each individual learner);
2. *What* will be modeled? – Different learner characteristics (goals, plans, knowledge, skills, preferences, emotional and mental states, personality traits, etc.);
3. *How* will the model be developed? – Strategies for acquiring and maintaining the model;
4. *Why* will the model be developed? – Motivations for modeling (develop a profile of the learner, support the customization of assistance and feedback, facilitate the interpretation and prediction of learner behavior, etc.).

Various kinds of learner-related information may be included in a student model. Jameson identified five categories of such information (Jameson 1995: 246):[193]

- *Personal characteristics*. Factual, objective information about the learner, such as his or her demographic and academic background, performance records, and previous interactions with the agent or system;
- *Position on dimensions*. Ratings of the learner on various general or specific dimensions, such as proficiency, preferences, personality, expectations, experiences, etc.;
- *Long-term cognitive states*. Knowledge, beliefs, and goals of the learner that are important across tasks and situations;

[193] See also the lists of user-related characteristics provided in Chapter 6.1 and in Section 12.2.1.

- *Short-term cognitive states and events*. Events and temporary states occurring in the context of the task in which the learner is currently engaged;
- *Observable states and events*. States and events that can be reliably detected by people, agents, or systems, such as learners' actions and their observable causes and outcomes.

In general, the contents of a particular student model will depend on what information the instructional module of the modeling agent or system requires and can use (Self 1990: 114).[194] So far, the focus has mostly been on modeling the student's *knowledge*, chiefly because information about other (e.g. affective or motivational) aspects of the student is difficult to obtain and interpret (Abdullah 2003: 17). Furthermore, the interest in the learner's knowledge reflects the common perception in the early days of intelligent tutoring systems that the purpose of student models is to support the remediation of learner errors (p. 115). This view encouraged the development of mechanisms for judging the correctness of student knowledge with respect to a body of pre-defined (expert) knowledge (p. 117, cf. Section 12.2.3.2). In contrast, Self argued that student modeling should be concerned with describing learners' subjective *beliefs* (rather than their objective knowledge) in their own terms (rather than judge them against pre-specified 'correct' knowledge) in order to support processes that challenge learners to reflect on the justifications and implications of their beliefs (p. 117). On the practical side, for many subject areas, including virtually every domain that is not part of mathematics and the natural and engineering sciences (the traditional domains of intelligent computer-based tutors), it is very difficult to identify and describe 'correct' vs. 'incorrect' knowledge (Self 1990: 118; Good & Berger 1998: 223).[195]

A number of taxonomies for classifying student models have been proposed in the literature. One taxonomy (Self 1988), which emphasizes the different functions of student models, has been widely cited (e.g. Self 1990: 117; Nwana 1991: 3; Brusilovsky 1994: 77f.; Schulmeister 1997: 184; Wilson & Villa 2002: 7f.; Abdullah 2003: 15f.):

- *Corrective*. The student model helps to determine necessary remediating actions by detecting differences between the 'correct' (expert) knowledge and the learner's current understanding of the subject matter.
- *Elaborative*. The student model is used to extend or refine the learner's incomplete current knowledge. Activities for elaboration may be selected based on a curriculum, a comparison of student and expert knowledge, an analysis of the contents of the student model, or a set of options generated for the learner to choose from.
- *Strategic*. The student model captures information about the learner, for example concerning the individual effectiveness of different teaching strategies, that enables the agent or system to make strategic changes to its operation, such as using a different tutoring approach or adapting the interaction style to the learner.
- *Diagnostic*. The student model contains information that can be inspected or queried to analyze the learner's knowledge state and to clarify what elements are incorrect or missing in his or her knowledge.

---

[194] In fact, it has been argued that student models in intelligent tutoring systems have often included information for which the rest of the system had no use (Self 1990; Beck et al. 1996; VanLehn et al. 2005), hence Self's warning "Don't diagnose what you can't treat" (Self 1990: 114).

[195] Philosophers of science actually have long maintained that there is no scientific knowledge that is final and invulnerable to falsification (Self 1990: 115).

- *Predictive*. The student model simulates the learner's behavior, which allows predictions about the effects of particular instructional actions on the learner.
- *Evaluative*. The student model contains information that enables the agent or system to assess the learner's level of achievement and/or its own success in teaching or assisting the learner.

Since the first reported student model in an intelligent tutoring system (Carbonell 1970), the problem of building an accurate, broad, and computationally efficient model of the individual learner has generated much interest in the ITS community and beyond. Many different techniques have been proposed for this purpose over the years. The most important of them will be reviewed in the next section. However, all of them have had to find solutions to a number of fundamental issues, which are discussed below:

- *Initialization*. What is the initial state of the student model? Student modeling has to start with some kind of initial model of the learner that contains assumptions about his or her background, preferences, goals, knowledge of domain concepts, and so on. Some of these assumptions may turn out to be correct while others may have to be revised during later updates to the student model. Eliciting a learner profile by means of explicitly requesting information from the learner is one way of obtaining an initial state for the student model (Self 1994a); activating learner stereotypes (Kay 2000; Tsiriga & Virvou 2003, cf. Section 12.2.1) or assuming a default learner profile are further possibilities (Self 1994a).
- *Bandwidth*. What is the amount and quality of the information about the student's activities and characteristics that is available to the student modeling process (VanLehn 1988: 58; Amižić et al. 2002)? At one extreme, student modeling may only have access to the learner's answer or result for the current task. Other approaches monitor the student's observable actions while performing a task as well as his or her intermediate and final results. Still others combine information about the student's observable behaviors and results with some kind of model of his or her cognitive processes (VanLehn 1988: 58f., cf. Section 12.2.3.5). In addition, information about individual characteristics of the learner, such as motivation, preferences, and personality, may be available in some cases (although, for the most part, it is not) (Abdullah 2003: 17). A history of the learner's actions, results, states, etc. can also be created.
- *Knowledge acquisition*. How are the elements and the structure of the student model and the necessary background knowledge[196] for student modeling acquired? In Section 12.1.1, two categories of knowledge acquisition methods were discussed: human knowledge engineering and machine learning. If permitted by the formalism used for modeling, combining both paradigms can facilitate the knowledge acquisition process and improve its outcomes. So far, manual encoding of student models and background knowledge has been predominant, but a growing number of systems build at least part of their structures and knowledge with the help of machine learning (Sison & Shimura 1998: 128). See the article by Sison and Shimura (1998) for a review of the use of machine learning techniques to induce student models and to develop background knowledge for student modeling. Machine learning for user modeling is reviewed by Webb et al. (2001).

[196] According to Sison and Shimura, the background knowledge of a student modeling agent or system includes 'correct' domain knowledge (cf. Section 12.1), 'faulty' student knowledge (cf. Section 12.2.3.2), (historical) knowledge about the individual learner, and stereotypical knowledge about categories of learners (Sison & Shimura 1998: 130f.).

- *Knowledge type*. What is the nature of the knowledge to be represented (Amižić et al. 2002)? Many systems have modeled procedural and/or declarative knowledge (cf. Table 17) of the subject area (domain) (VanLehn 1988: 59ff.; Schulmeister 1997: 182f.), typically using the same formalism for both the domain knowledge and the student knowledge (VanLehn 1988: 62; Abdullah 2003: 17) and imposing a hierarchy or some other organization on the knowledge representation (Wilson & Villa 2002: 9). Procedural knowledge comprises a set of procedures that specify how to perform particular tasks, often in the form of production rules (cf. Section 12.1.2.4). Procedural models can be executed by an interpreter to simulate the problem-solving behavior of the learner or the domain expert (Brusilovsky 1994: 72; Abdullah 2003: 16). Declarative knowledge consists of facts and rules organized in a way that enables the system to reason with them in order to verify the consistency of facts, derive new facts, etc. Quite a number of intelligent tutoring systems have used semantic networks (cf. Section 12.1.2.2) for this purpose (Wilson & Villa 2002: 24). The corresponding declarative student model is often used to capture the learner's level of mastery of the domain facts and rules.
- *Student-expert difference*. How do the models of student knowledge and expert knowledge differ (Amižić et al. 2002)? Some items which are in the expert model may be missing from the student model. Conversely, the student model may contain items that are incorrect (or 'buggy') and therefore are not part of the expert model (VanLehn 1988: 62). Other items in the student model may be correct, but they may not be represented in the expert model (Friedland 2001: 58). It should be noted that while the expert paradigm (Brusilovsky 1994: 81ff.), which places the system in the role of an omniscient tutor, has dominated research on student modeling for a long time, this approach has turned out to be infeasible from an implementation point of view and undesirable in light of the currently prominent constructivist theories of learning (cf. Chapter 2.2.3) (Self 1990: 115–118; Brusilovsky 1994: 82). More recent alternative approaches to student modeling have abandoned the idea of the omniscient tutor and have turned to greater learner involvement, sufficiently accurate representations of the learner's beliefs (rather than models of his or her correct and incorrect knowledge), open student models (see below), and the role of the system as an assistant or collaborator (Self 1990: 118; Brusilovsky 1994: 82f.).
- *Grain size*. At what level of detail is the student's knowledge of the subject matter being modeled? The granularity of student models typically ranges between the two extremes of noting that the learner does or does not know the domain on the one hand and keeping track of every action and state of the learner on the other. Most systems model both the domain and the student's knowledge at the same level of granularity (Beck et al. 1996).
- *Completeness and precision*. How accurate is the learner-related information being modeled? Does the model cover all (relevant) aspects of the student? In general, building a complete and fully accurate student model has proven to be an intractable problem (Self 1990), given the potentially large number of factors involved, their complex and numerous relationships, and the difficulty to detect, track, and model both the factors and their relationships over time. In fact, even constructing a partial model with reduced accuracy is generally a non-trivial task (Abdullah 2003: 38). Given that student models are incomplete and imprecise at best, a number of researchers have argued that having no student model at all would be better than having to live with one that is less than optimal. This view has led these researchers (e.g. Gugerty 1997) to give up the whole endeavor of student modeling as infeasible and even unnecessary (Abdullah 2003: 38). However, it appears unreasonable to assume that intelligent tutoring systems or pedagogical software agents would be able to assist the individual learner effectively if they knew nothing about the student and his or her

development. Besides, as argued above, in practice, both completeness and full precision may actually matter less than performance since the results of student modeling may have to be available on time while the learner is interacting with the agent or system.

- *Level of analysis*. At what level(s) are the relationships between observable learner behaviors and the background knowledge of a student modeling system analyzed? Sison and Shimura described three levels of (error) analysis (Sison & Shimura 1998: 132): the first level, the *behavior level*, identifies syntactic mismatches between the learner's behaviors and a set of pre-specified desired behaviors. These discrepancies are interpreted as behavioral errors. The next higher level, the *knowledge level*, uses the results of the behavior level to detect both inconsistent or incorrect and missing or incomplete knowledge that may be responsible for behavior-level errors.[197] At the highest level, the *learning level*, such knowledge errors are related to corresponding, analogous, or otherwise relevant items of knowledge. For example, misconceptions (incorrect or inconsistent knowledge, cf. Section 12.2.3.2) may be linked to items from which they possibly arose, e.g. by way of overgeneralization. In the hierarchy of the three levels (behavior, knowledge, and learning), each higher level accounts for errors at the next lower level.
- *Time frame*. What is the temporal scope of the student model? *Long-term* student models accumulate information about learners over a series of interactions, including their growing level of knowledge, recurring errors, etc. In contrast, *short-term* models capture the learner's conceptions and misconceptions at a particular point in time, which are derived by reasoning about the learner's behavior in the current context. Ideally, systems providing individualized learner support should have both types of models (Kabassi & Virvou 2003: 61). From another point of view, the distinction between short-term and long-term modeling can also be made in terms of local vs. global modeling. *Local modeling* traces the learner's current performance using online monitoring and error recognition. *Global modeling* uses the trace data generated by local modeling to develop a persistent model of the learner's knowledge across interactions (Johnson & Taatgen 2005: 433).
- *Inspectability*. Can the learner being modeled inspect, and possibly even modify, the contents of his or her own student model? Most student modeling approaches have used internal student models that are manipulated by the system alone (Self 1990: 120). However, opening student models to learners and involving them in their construction may increase the accuracy of the resulting models and provide additional educational advantages, such as increased learner reflection and learner control (Self 1990: 120; Abdullah 2003: 40, cf. Section 12.2.3.7).
- *Sharability*. Is it possible for different agents or systems to share their student models? The shared student modeling data could be used to quickly adapt to the characteristics of a new learner, and also inform the development of better learner-adaptive systems (Gertner 1999). The sharing of student models would benefit from a common, domain-independent standard for student modeling, which allows to clearly specify what learner-related aspects are being modeled at what level of detail (grain size) and what the measurements used mean (Gertner 1999). Such a standard is not yet available.

[197] Sison and Shimura also allowed for the possibility of behavioral errors being caused by slips in the learner's performance (Sison & Shimura 1998: 132).

12.2.3 Approaches to Student Modeling

This section reviews a number of approaches to student modeling that have been described in the literature, chiefly in the context of intelligent tutoring systems. The advantages and shortcomings of each approach are discussed.

12.2.3.1 Student Modeling Using Scales and Tallies

The simplest way to model a learner's level of knowledge about a domain is to measure it on a scale (Brusilovsky 1994: 72). Scalar models indicate the learner's knowledge level in terms of a number on a scale between a minimum and a maximum value representing the lowest and the highest possible estimate, respectively (p. 72). Scales can be maintained both for the learner's knowledge as a whole and for different components of his or her knowledge.

Another simple student modeling technique that has been widely used in computer-assisted instruction involves keeping a record, or tally, of the correct and incorrect answers given by the learner (Rickel 1989). Some limited forms of adaptation are possible with this technique, in particular taking a different branch of instruction if the learner's current answer is (in-) correct or if he or she has given a certain number of correct (or incorrect) answers (Rickel 1989, cf. Chapter 2.3.2).

The problem of both the scalar and the tally models is that they can only provide a very coarse representation of the learner's knowledge state, which may be sufficient for calculating a grade for the learner's performance but is clearly unsuitable for providing truly individually adaptive instruction (Rickel 1989). There is no representation of what items the learner knows and does not know and where he or she may have incorrect or inconsistent conceptions about the domain. The overlay model and its extensions were developed to address these shortcomings (Brusilovsky 1994: 72).

12.2.3.2 Overlay Student Modeling

For much of their history, intelligent tutoring systems have used student modeling mainly to capture the system's evolving view of the learner's developing knowledge or understanding of the subject matter (cf. Section 12.2.2). An intelligent tutor analyzes the learner's activities and responses, referring to its model of domain knowledge, in order to identify gaps and errors in the learner's knowledge that call for an intervention by the tutor (Self 1999: 362). The gaps and errors correspond to underlying missing conceptions and misconceptions of the learner, respectively (VanLehn 1988: 62). Both missing conceptions and misconceptions are defined in terms of differences between the domain (expert) model (cf. Section 12.1) and the student model. The term *missing conception* refers to an item of knowledge (a concept, fact, principle, procedure, schema, strategy, or topic (Beck et al. 1996; Sison & Shimura 1998: 132)) that is present in the expert model but not in the student model (VanLehn 1988: 62). *Misconceptions* are items of knowledge which are present in the student model but not in the expert model, and which are therefore assumed to be incorrect or inconsistent conceptions (p. 62).[198]

[198] A student may actually possess correct knowledge which an expert (model) does not have (Friedland 2001: 58). However, traditionally, intelligent tutoring systems have implicitly claimed for themselves the position of the sole authority in their domain of expertise (Self 1990: 115).

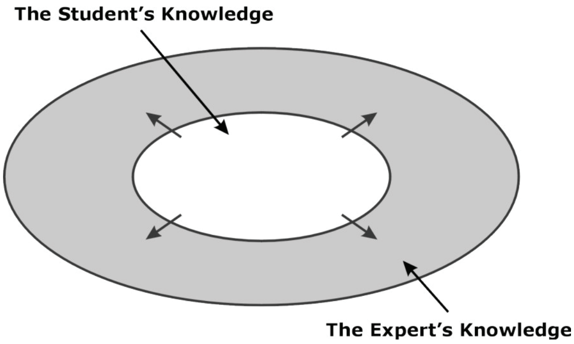

Figure 63. The schema of the overlay student model. Adapted from Friedland (2001: 58, Figure 4.4).

If only missing conceptions are considered, the student model becomes a subset of the expert model (Wenger 1987: 137; VanLehn 1988: 62). In other words, the student model is structured in the same way as the expert model and contains the same items (though not all of them) (Wilson & Villa 2002: 9; Abdullah 2003: 17). The student is regarded as an 'expert-in-training' who knows part of what the expert knows, but not yet everything. A comparison of the learner's and the expert's knowledge (referred to as 'overlay') yields differences (missing conceptions) that are addressed by exposing the learner to material which increases his or her knowledge so that it eventually approximates the knowledge of the domain expert (Beck et al. 1996; Friedland 2001: 58). Figure 63 shows this *overlay model* (Carr & Goldstein 1977), which has been widely adopted for student modeling in intelligent tutors and other user-adaptive systems for education (VanLehn 1988: 62; Sison & Shimura 1998: 133; Wilson & Villa 2002: 9) because it is conceptually simple and easy to implement while still effective as a model of the learner's knowledge in applications that involve teaching the expert knowledge to the learner (Baffes 1994: 5; Holt et al. 1994: 6f.).

One problem of the overlay model is that only missing conceptions (lack of knowledge) can be represented (VanLehn 1988: 62; Abdullah 2003: 18). Furthermore, no distinction is made between items of knowledge that the learner has seen but not yet mastered and items which the learner cannot possibly know due to lack of exposure to these items (Friedland 2001: 58; Abdullah 2003: 18; [INT 146]). The latter problem has been addressed by *differential student models* (Burton & Brown 1982), which extend the overlay model as shown in Figure 64. The student's knowledge is modeled as a subset of the knowledge that has been presented to the learner, and both are embedded in the expert's knowledge (Friedland 2001: 58; [INT 147]). The goal of the system is to bring the learner closer to expert-level proficiency in two steps, by first increasing his or her knowledge of items to which he or she has already been exposed (using an overlay model) and then introducing him or her to new material (Friedland 2001: 58). While an improvement over simple overlay models, differential student models are still unable to deal with misconceptions or correct student knowledge that is not part of the expert knowledge modeled in the system (p. 58).

To handle misconceptions, the expert knowledge in overlay models has been augmented with a representation of faulty knowledge, which is called a *bug library* (Baffes 1994: 5; Holt et al. 1994: 8f.).[199] This library is a database of common misconceptions (also referred to as

[199] The terms *buggy knowledge* and *bug catalogue* are also used (Abdullah 2003: 20).

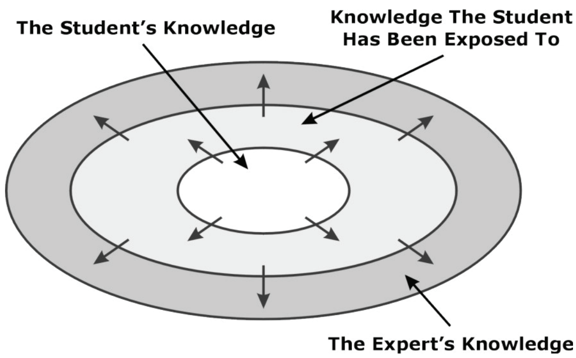

Figure 64. The schema of the differential student model. Adapted from Friedland (2001: 58, Figure 4.5).

bugs[200]) that are held by students (Baffes 1994: 5; Sison & Shimura 1998: 131; Friedland 2001: 58). Given a bug library, diagnosing the learner's misconceptions involves matching his or her observed behaviors against the catalogue of pre-defined bugs (VanLehn 1988: 62; Baffes 1994: 5). Bug libraries can simply list common bugs that have been pre-constructed using sources including the educational literature, an analysis of learners' behaviors (in particular errors), and a learning theory for the domain (*bug enumeration*), or they may provide a collection of bug parts from which bugs are dynamically generated while diagnosing the student (*bug generation*) (VanLehn 1988: 62f.; Baffes 1994: 5; Sison & Shimura 1998: 131; [INT 148]). Student models involving bug libraries are instances of the *perturbation* or *buggy student model* (Burton 1982), which is illustrated in Figure 65. By applying an overlay model to the augmented model consisting of expert knowledge and bug library, the system seeks to both extend the student's knowledge and reduce the number of detected learner misconceptions (Holt et al. 1994: 8f.; Friedland 2001: 58). Some differences between the student model and the expert model are regarded as less important than others, including cases where the learner has correct knowledge which the system does not have (p. 58).

Diagnosis and remediation of student misconceptions are hallmark capabilities of intelligent tutoring systems, which were the concern of much work in the early days of the field (Corbett et al. 1997: 858). At first, student models involving bug libraries seemed to be appropriate for this task. However, while an important extension of the overlay student model, the bug library approach is faced with a number of serious issues. First, the construction of a bug library is costly in terms of time and effort (Baffes 1994: 6). Second, a complete bug model for a given area of expertise is effectively unattainable because no matter how much effort may be put into its construction, the resulting bug library will quite likely miss a substantial number of student misconceptions (Baffes 1994: 6; Sison & Shimura 1998: 131). Systems that can extend or build bug libraries are still very rare (Sison & Shimura 1998: 131), although some researchers have investigated the use of machine learning techniques (cf. Section 12.1.1.2) for the construction of bug libraries (e.g. Baffes 1994; Baffes & Mooney 1996). Third, bug libraries are static resources and therefore unable to model learner behaviors that were not anticipated during their construction (Baffes 1994: 6). Fourth, the issues of bug interaction and bug migration have to be considered. *Bug interaction* means that an observed student error may be the result of a combination of underlying misconceptions (Stevens et al. 1982; Rickel 1989). The problem of

[200] Different labels for items of faulty knowledge can be found in the literature, including *bug*, *error*, *mal rule*, and *misconception*. Authors tend to use these labels interchangeably and with different meanings (Abdullah 2003: 20).

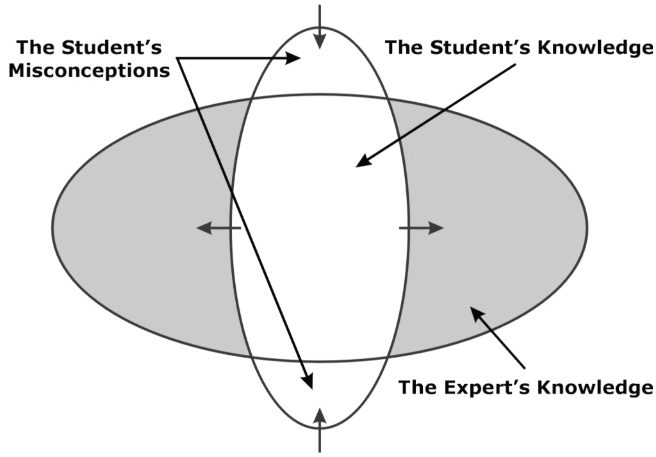

Figure 65. The schema of the perturbation student model. Adapted from Friedland (2001: 58, Figure 4.6) and ([INT 148]).

bug migration occurs when a bug apparently changes ('migrates') over time into another that is different but related (Brown & VanLehn 1980: 410; VanLehn 1982; Abdullah 2003: 21). Fifth, it may be necessary to construct different bug libraries for different groups of learners (Payne & Squibb 1990; Sison & Shimura 1998: 131). Sixth, enduring misconceptions are not as frequent as originally thought (VanLehn 1990; Sleeman et al. 1991; Corbett et al. 1997: 858). Seventh, empirical research has questioned both the instructional effectiveness and the cost-benefit ratio of remedial feedback based on bug modeling (Sleeman et al. 1989; McKendree 1990; Corbett et al. 1997: 858).

12.2.3.3 Student Modeling Using Genetic Graph Models

Overlay models provide a snapshot of the present state of the learner's knowledge compared to the knowledge of a domain expert, but they cannot account for the evolution of the learner's knowledge over time during the learning process (Abdullah 2003: 22). To model how learners progress from simple to complex and from particular to general knowledge (Brusilovsky 1994: 72), *genetic graph models* (Goldstein 1982; Wasson 1985; Brecht & Jones 1988) have been proposed, which define possible paths by which learners develop knowledge of the subject area (Holt et al. 1994: 7). A genetic graph is a kind of semantic network (p. 7, cf. Section 12.1.2.2) in which nodes correspond to items of (originally procedural) knowledge at different levels of complexity or generality (Brusilovsky 1992) and links between nodes represent evolutionary relationships between the connected items, including analogy, generalization, refinement, and deviation (Wenger 1987: 141), which can be regarded as domain-independent mechanisms for learning a given item from the one linked to it (Wenger 1987: 403; Vassileva & Wasson 1996). For definitions of the evolutionary links used in genetic graphs, see Wasson (1985: 32).

Genetic graphs extend the overlay-based student models discussed in the previous section (Wenger 1987: 140f.; Brusilovsky 1994: 72). The learner's knowledge and his or her learning behavior (represented by his or her individual learning path) are represented as an overlay on the nodes and the links of the genetic graph, respectively, including correct and incorrect pieces of knowledge, which are connected by deviation links (Wasson 1985: 33 + 45; Wenger 1987: 141f.). Thus not only the current state of learners' knowledge can be modeled but also their progress in acquiring different aspects of expertise (Holt et al. 1994: 7f.; Abdullah 2003: 22). Information about both the learner's path and possible extensions to it can be used to flexibly adapt future instructional interventions to the learner's progress (Wasson 1985: 39 + 41; Vassileva & Wasson 1996).

While genetic graph models incorporate some degree of learner orientation into the student model (Wasson 1985: 33; Wenger 1987: 140), they are still based on the overlay approach and therefore define the knowledge of the learner as a subset of the expert's knowledge (Abdullah 2003: 22). Furthermore, genetic graphs are static models consisting of pre-defined nodes and links that cannot be changed at run-time (Wasson 1985: 52; Wenger 1987: 144; Greve 1990). For complex domains, pre-constructing a complete static graph should be very difficult, if not impossible (Wenger 1987: 144). As a result, there may be learning paths that the static model is unable to account for. While Goldstein hinted at the possibility of building genetic graphs with dynamically generated links and nodes, the feasibility of this idea for complex domains remains doubtful (p. 144). A theoretical issue is that the evolutionary relations in a genetic graph do not provide a full account of the complex learning processes which they represent (p. 144). In addition, learning is taken to involve only local additions and refinements of knowledge; major reconceptualizations are not possible (p. 144).

12.2.3.4 Probabilistic Student Modeling

There is typically a significant gap between the nature and amount of the information that an intelligent system can feasibly and reliably collect about the learner and the inferences that have to be made from that information about the learner's beliefs, abilities, goals, motivations, and future behavior (Jameson 1995: 193). In other words, student modeling (and user modeling in general) normally has to deal with considerable *uncertainty* with respect to the conclusions that have to be drawn from the comparatively meager and haphazard evidence available (p. 193).

One important numerical approach to uncertainty management in student modeling makes use of Bayesian networks[201] (cf. Section 12.2.1) to model the uncertainty of determining the learner's knowledge state and plans (Martin & VanLehn 1995; VanLehn & Martin 1997; Conati et al. 1997; Mayo & Mitrovic 2001; Conati et al. 2002). In particular, Bayesian networks have been proposed as a solution to the so-called *assignment of credit problem* in student modeling (VanLehn 1988). That is, if an observed action of the learner has multiple possible explanations with respect to the learner's knowledge and/or plan, how much credit should be assigned to each of them? In probabilistic terms, how likely is each explanation? Which one is the most likely? (Conati et al. 2002: 372).

In student modeling, Bayesian networks have been mainly used to infer from the learner's observable actions how well he or she knows the domain and what his or her actions are trying to accomplish, in other words, for *student assessment*[202] (Martin & VanLehn 1995) and for *plan recognition* (cf. Section 12.2.1). A great deal of inherent uncertainty is involved in both tasks when learners can reason along various lines and do not have to make explicit all the steps in their reasoning (Conati et al. 2002: 371). A third application of Bayesian networks in student modeling is to provide support for the *prediction* of the learner's behavior while he or she is

[201] The *Dempster-Shafer theory of evidence* (Dempster 1968; Shafer 1976) and *fuzzy logic* (cf. Section 12.1.2.1) are two further important frameworks for numerical uncertainty management in user and student modeling. See Jameson (1995) for a review.

[202] Another term for student assessment is *knowledge tracing* (Conati et al. 2002: 371). A discussion of knowledge tracing (which also employs Bayesian techniques) is provided in the following section on cognitive student modeling.

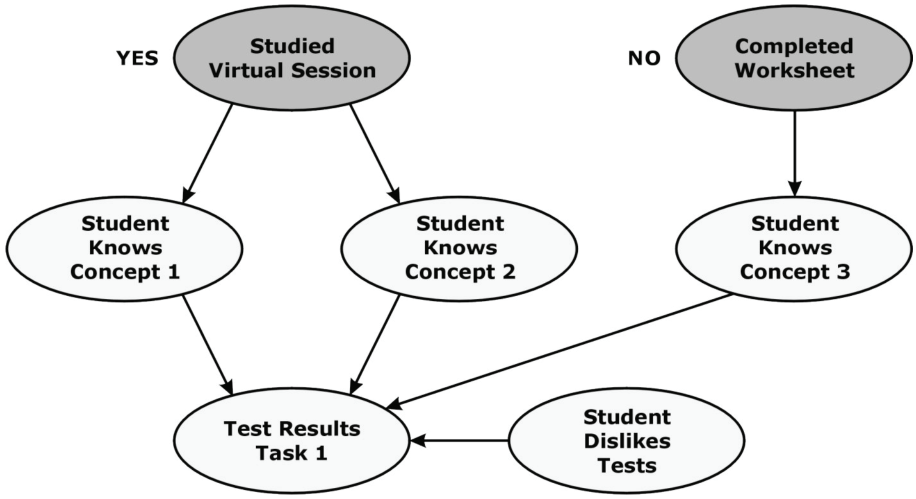

Figure 66. A simple Bayesian student model. Adapted from Mayo and Mitrovic (2001: 126, Figure 2).

engaged in problem solving. If the learner's inferences and actions in a given problem-solving context can be predicted, it becomes easier to assist him or her in overcoming an impasse (p. 373).

A simple example (adapted from Mayo & Mitrovic 2001) will suffice to illustrate how the relationships between observed actions of the learner, his or her internal states, and a set of outcomes can be modeled with Bayesian networks (p. 126). Figure 66 shows a small Bayesian network that can be used to predict the learner's performance on a particular task in a test, given how likely it is that he or she knows three different concepts (which, in turn, depends on the probability that he or she has performed certain learning activities) and the likelihood that the learner is not fond of tests in general. The two peripheral nodes in dark grey, "Studied Virtual Session" and "Completed Worksheet," represent variables whose values were observed in the learner's behavior and which therefore could be instantiated (set with certainty to) YES and NO, respectively (cf. Section) (p. 126). As a result of the instantiation, the probability distributions of all non-instantiated nodes in the network are updated to integrate these observations, which leads to updated probabilistic beliefs about the non-instantiated variables (cf. Section 12.2.1).

Bayesian networks provide a consistent, robust, flexible, and theoretically solid mechanism for managing uncertainty in student modeling (Myllymäki 2005). *Consistency* means that outputs computed using a Bayesian network model are always unambiguous and that alternative ways of computing an output from a given input using the network will produce the same result. The performance of a Bayesian network is *robust*, i.e. not severely affected by minor changes to the network structure. Bayesian networks are *flexible* structures because the same network model can be used for a variety of tasks, including student assessment, recognition of the student's plans, and prediction of learner steps (see above) (Myllymäki 2005). Furthermore, Bayesian networks have a *solid theoretical basis* in probability theory, graph theory, computer science, and statistics (Ben-Gal 2007). Bayesian networks are also *transparent* because their structure, variables, and probabilities can be inspected and understood, not only by their developers but also by domain experts, in contrast to artificial neural networks (cf. Chapter 11.1.1.2 and Section 12.2.1), whose internals are not accessible to the observer (Myllymäki 2005). Combined with well-understood formal mechanisms for computing probabilities and making inferences, this ensures that a given Bayesian student model can be used to explain its own outputs (Gertner 1999). Finally, Bayesian networks can incorporate information from *various sources*. For example, a domain expert bootstraps the network, the rest is automatically learned from data, and the expert makes the final adjustments (Mayo & Mitrovic 2001: 127).

Despite these advantages, the construction and performance of Bayesian student models pose significant challenges for developers. Their task is to build the structure of a Bayesian network and to define the probability distributions of its nodes in such a way that the final network has sufficient accuracy and coverage and can be used efficiently in processing. In short, problems related to the acquisition and computational complexity of Bayesian networks have to be solved (Jameson 1995: 239–243).

In general, the quality and trustworthiness of the probabilities computed from a Bayesian student model depend on the extent to which the prior and conditional probabilities specified for the nodes of the network are reliable (Gertner 1999). Determining the prior and conditional probabilities for a given domain is a difficult knowledge engineering task (cf. Section 12.1.1.1). Probabilities have been obtained from a variety of sources, including domain experts, empirical data, the research literature, machine learning (see below), and (quite often) more or less informed guessing (Jameson 1995: 197; Gertner 1999; Conati et al. 2002: 388).

Depending on how they develop the structure and the prior and conditional probabilities of Bayesian student models, current methods have been classified intro expert-centric, efficiency-centric, and data-centric approaches (Mayo & Mitrovic 2001: 127–131; González et al. 2006). *Expert-centric* approaches develop all or most of the structure and probabilities of a Bayesian network model based on an analysis of the domain provided by a human expert, imposing no structural restrictions on the network in order to model the structure of the domain as faithfully as possible (Mayo & Mitrovic 2001: 127). The resulting networks often have a large number of variables, many of them specifying unobservable internal states of the learner. It is difficult to obtain probability distributions for these variables without data (p. 127f.). In addition, dynamic adaptation of the structure and/or parameters of the network to the learner is not possible (p. 129). Finally, the size and complexity of expert-centric network models needed for realistic domains have been found to cause performance problems (see below).

Concerns about the lack of efficiency of Bayesian student models have motivated *efficiency-centric* approaches, which try to make building these networks and processing with them more efficient by specifying them partially (rather than completely) or by restricting their structure (Mayo & Mitrovic 2001: 127; González et al. 2006). However, restrictions introduced to reduce complexity (and thereby increase efficiency) may lead to an oversimplified model that possibly even includes incorrect assumptions about the domain (Mayo & Mitrovic 2001: 128; González et al. 2006).

A third possibility is to learn Bayesian student models from data which has been collected from a representative group of learners by exposing them to a reduced version of the final system that selects its instructional actions randomly (Mayo & Mitrovic 2001: 132f.). The next step is to apply machine learning techniques (cf. Section 12.1.1.2) to the data in order to induce the structure and probabilities of a Bayesian network that can predict student performance (p. 132f.). This model, in turn, is gradually adapted to the current learner over time (Mayo 2001: 110; Mayo & Mitrovic 2001: 144). The details of this *data-centric* approach to Bayesian student modeling have been described in the literature (Mayo 2001; Mayo & Mitrovic 2001). One advantage of data-centric Bayesian student models is that their predictive performance can be readily tested using unseen data that was not part of their training data set (Mayo & Mitrovic 2001: 130). Furthermore, Mayo and Mitrovic pointed out that data-centric networks usually contain a much smaller set of variables than expert-centric models because they only represent the relationships between observable variables to predict learner performance while expert-centric networks model both observable and hidden variables (Mayo & Mitrovic 2001: 128 +

130). One problematic aspect of the data-centric approach is that the performance of the resulting student models depends on the amount and quality of the training data that can be obtained for the machine learning process (cf. Section 12.1.1.2), although human involvement is possible both before and after the learning stage (Mayo & Mitrovic 2001: 133). Furthermore, the purpose of the Bayesian student model seems to be to inform "pedagogical action selection" (i.e. selection of problems, error messages, topics, etc.) (p. 138) rather than assessment of the learner's knowledge: "The output of the Bayesian network (which is a prediction about how the student will (…) behave in a given context) is the input to a decision-theoretic process for tutorial action selection" (Mayo 2001: 9). Finally, modeling only observable variables makes it easier to acquire the network, but reasoning about unobservable aspects of the tutorial state, including the learner's knowledge, is not possible with the resulting model (Murray et al. 2004).

Concerning computational complexity, a general problem of large Bayesian networks is that their online evaluation (i.e. while the learner is working with the system) is computationally expensive, even to the point of intractability (Mayo & Mitrovic 2001: 127). In fact, performing exact inference in large, complex Bayesian networks is known to be NP-hard (i.e. no exact solution can be found in polynomial time) and hence computationally intractable in general (Jameson 1995: 242; Russell & Norvig 2003: 511; Ben-Gal 2007). Inference from a Bayesian network involves belief propagation, i.e. if new information about a subset of the variables in the network is available, it is used to update the uncertain beliefs about the other variables (Laskey 2008, cf. Section 12.2.1). In the worst case, the number of operations involved in the probability propagation algorithms grows exponentially with the number of the variables in the network (Ben-Gal 2007). While there is no solution to the general problem of Bayesian inference, tractable algorithms exist for singly connected networks with at most one undirected path between any two of their nodes (Jameson 1995: 242), for some types of low-connectivity networks, and for networks suitable for approximate inference techniques (Russell & Norvig 2003: 504–519; Laskey 2008).

In sum, Bayesian networks are a versatile probabilistic framework for processing uncertain information, which has a solid and well-understood theoretical foundation. Not surprisingly, these networks have been widely adopted to manage uncertainty in student modeling. However, developers considering the use of Bayesian networks need to be prepared to address problems related to the acquisition of the structure of and probabilities for their networks as well as issues involved in the management of the computational complexity of Bayesian models, which can be considerable in realistic domains.

12.2.3.5 Cognitive Student Modeling

Cognitive approaches to student modeling have been quite influential in the field of intelligent tutoring systems over the last twenty years. Cognitive tutors (Anderson et al. 1995; Koedinger et al. 1997; Koedinger 2001) are intelligent computer-based tutors which are characterized by their use of a psychological model of the cognitive processes that underlie (near-) successful performance of learners in problem solving (Koedinger et al. 1997: 32). The *Adaptive Control of Thought(-Rational) (ACT(-R)) theory* of learning and performance (Anderson 1983; Anderson 1993; Anderson & Lebiere 1998; Anderson et al. 2004) provides the theoretical basis of this approach (Koedinger 2001: 147). The cognitive model of student problem solving built for the tutor consists of production rules (cf. Section 12.1.2.4) that represent (alternative) steps to solve problems in the domain, and mis-steps corresponding to typical learner misconceptions (cf. Section 12.2.3.2) (Koedinger et al. 1997: 32; Koedinger 2001: 149). At each point in the problem-solving process, the model can be used to dynamically generate a set of possible solution paths and to suggest correct and incorrect steps that could be performed by students at that stage (Anderson et al. 1995: 171; Amižić et al. 2002; Mitrovic et al. 2003). The cognitive model is involved in two student modeling techniques that trace the learner's problem solving and knowledge development, respectively.

Model tracing performs a step-by-step analysis of the learner's problem-solving activities, matching each student action to the actions that would be generated by those production rules in the cognitive model that are applicable in the current situation, in order to identify both *on-path actions* (i.e. legitimate actions, which lie on one of the solution paths generated by the cognitive model) and *off-path actions* (i.e. learner actions that are not on one of the correct paths and therefore assumed to result from misconceptions) (Anderson et al. 1995: 171; Corbett et al. 1997: 853; Johnson & Taatgen 2005: 433). Learner actions matching either one of the buggy production rules (which embody common learner misconceptions) or none of the production rules in the cognitive model are interpreted as errors and may trigger tutorial responses from the system (Mitrovic et al. 2003). While the learner is on-path, the tutor silently performs its tracing activities in the background and does not intervene (Anderson et al. 1995: 171; Koedinger et al. 1997: 32). However, this silent monitoring allows the tutor to keep track of the state of the learner's problem solving and to provide the student with timely, individualized and context-sensitive assistance (feedback on errors and problem-solving hints) (Koedinger et al. 1997: 32; Koedinger 2001: 157f.).

Knowledge tracing (Corbett & Anderson 1994) keeps track of how the knowledge of a given learner increases as he or she solves problem after problem, by maintaining a student-specific model of probabilistic estimates of the learner's mastery of each production rule in the generic cognitive model.[203] Based on these estimates, problem selection and pacing may be tuned to the individual learner (Koedinger 2001: 149). The estimates can also be made visible to the student or teacher, for example as a graphical scale (Koedinger et al. 1997: 36). To obtain the probabilistic estimates for the production rules, a Bayesian procedure (cf. Section 12.2.1) is employed, which was described by Corbett and Bhatnagar (1997). Each production rule is in one of two states: learned or unlearned. The transition from unlearned to learned (but not backwards) may take place before the rule can be applied for the first time (initial learning) or whenever it is possible to use the rule during problem solving (acquisition). Because slips and lucky guesses will occur in the student's use of learned and unlearned rules, respectively, knowledge tracing maintains for each production rule an estimate of its probability to be in the learned state, which is updated at each point where the rule is applicable in the problem-solving process, depending on whether the learner has taken a correct or an incorrect action. The Bayesian procedure for computing the probability uses a set of learning and performance parameters whose estimates are obtained empirically for each rule (Corbett & Bhatnagar 1997).

One major advantage of cognitive tutors is that they use an explicit model of the procedural knowledge which learners have to acquire in order to be able to solve problems in the domain (Mitrovic et al. 2003). Cognitive tutors evaluate each problem-solving step taken by the student and can respond appropriately. Feedback is typically provided immediately but can also be given on demand (Mitrovic et al. 2003). Furthermore, cognitive tutors can generate context-sensitive strategic hints on the next problem-solving step (when requested by the learner or when the number of his or her mistakes exceeds a pre-set threshold) with the help of the set of production rules in their cognitive model (Koedinger 2001: 157f.; Mitrovic et al. 2003).

[203] In other words, the student's knowledge is modeled as an overlay on the ideal knowledge (Corbett & Bhatnagar 1997, cf. Section 12.2.3.2).

Unrecognized problem-solving steps of the learner (i.e. steps that fail to trigger a correct or a buggy production rule) are assumed to be incorrect. In such cases, the tutor can indicate the deviant step, although it cannot offer an explanation of why that step is wrong (Mitrovic et al. 2003). However, in so doing, a cognitive tutor may also misjudge correct problem-solving steps if it lacks the relevant correct production rules (Mitrovic et al. 2003). In general, the "guilty until proven innocent" default behavior of cognitive tutors when analyzing student actions can lead to the rejection of correct solutions if the solution paths that can be generated by the cognitive model are unable to account for them (Mitrovic et al. 2003). In fact, tutors based on the model tracing approach have been criticized for their rigidity because they deny the learner the freedom to explore approaches that are outside their own pre-defined set of problem-solving strategies (VanLehn et al. 2000; Mitrovic et al. 2003). Another point of criticism concerns the high degree of scaffolding which these tutors provide for error detection and remediation: they identify every problem-solving error for the learner and generate an appropriate remedial hint, which, however, does not help students to develop these error handling capabilities themselves. In contrast, human tutors offer substantial scaffolding only at the beginning and then gradually fade as the learner's proficiency increases (VanLehn et al. 2000, cf. Chapter 2.2.3). Therefore, VanLehn et al. suggested that model tracing tutors should be equipped with the ability to fade their scaffolding. Furthermore, they should gradually increase the learner's freedom in problem solving by easing their restrictions on how to correctly solve problems in the domain (VanLehn et al. 2000).

A number of critics have also found fault with model tracing tutors because of their lack of support for deep learning (VanLehn et al. 2000). Some of the criticisms in this area were discussed by VanLehn et al. (2000) and are summarized in the following. First, learners are not encouraged to reflect on the hints provided by the tutor and can arrive at a correct action simply through trial and error. In so doing, they might learn how to solve problems, but only at a shallow level, without really understanding what they are doing. Second, students are not required to explain their problem-solving actions and therefore may not learn the terminology of the domain and how to use it properly. Third, model tracing tutors request learners to develop detailed solutions to specific problems. As a result, learners may fail to see the "big picture," i.e. the general approach that enables them to solve a whole class of similar problems. Finally, students learning a quantitative skill are not invited to take a *qualitative*, semantic viewpoint on the problems they are solving and thus do not develop the ability to tackle related qualitative problems and to determine whether a given quantitative problem makes sense. VanLehn et al. suggested that these shortcomings could be addressed by combining the problem-solving expertise of a model tracing tutor with knowledge about principles and rationales; domain language; abstract, basic approaches; and qualitative rules of inference. They further argued that "if [model tracing tutors] are to become more effective at encouraging deep learning, they must use natural language" in instructional interactions between the tutor and the learner (cf. Section 12.3.9) that help learners to acquire the kinds of knowledge mentioned above (VanLehn et al. 2000).

12.2.3.6 Constraint-Based Student Modeling

Constraint-based modeling (CBM) (Ohlsson 1992; Ohlsson 1994; Mitrovic & Ohlsson 1999; Mitrovic et al. 2001; Mitrovic et al. 2003; Ohlsson & Mitrovic 2006) has been developed as a solution to the intractable problem of creating and maintaining a complete and precise model of the student's knowledge (Mitrovic et al. 2003, cf. Section 12.2.2). The central idea of CBM is to model knowledge about a given domain in terms of an inventory of *constraints* on correct solutions in that domain (Mitrovic & Ohlsson 1999: 239). These constraints divide the space of possible solutions into correct solutions that satisfy and incorrect solutions that violate the constraints. Put differently, given the set of all possible solutions, the constraints select those solutions from this set which are correct (p. 239). For each constraint, the corresponding set of incorrect solutions consists of all solutions violating that constraint (p. 239). The learner's knowledge of the domain can be modeled in terms of the constraints which are violated and the ones which are not violated by the learner during problem solving (Mitrovic & Ohlsson 1999: 239; Ohlsson & Mitrovic 2006: 16). Instead of modeling the learner's *actions*, the student model captures the *effects* of his or her actions (Mitrovic et al. 2001). Diagnostic information is assumed to be obtainable from the learner's current situation (i.e. the problem state) rather than the sequence of actions by which he or she reached that state (Mitrovic et al. 2001).

The basic unit of knowledge in constraint-based models is the *state constraint*, an ordered pair $<C_r, C_s>$ consisting of the *relevance condition* C_r and the *satisfaction condition* C_s (Mitrovic & Ohlsson 1999: 239). C_r represents "the class of problem states for which the constraint is relevant," whereas C_s is "the class of (relevant) states in which the constraint is satisfied" (p. 239). Both C_r and C_s are described in terms of features or properties (p. 239). A state constraint can thus be read as: *If the constraint is relevant (the properties C_r hold), then it has to be satisfied (the properties C_s have to hold as well). Otherwise, something is wrong* (p. 239). "In other words, if the student solution falls into the state defined by the relevance condition, it must also be in the state defined by the satisfaction condition in order to be correct" (Mitrovic et al. 2006). To give a concrete example: *If the code of an if-condition statement contains an opening curly brace ('{') (= C_r), there has to be a closing curly brace ('}') as well (= C_s). Otherwise, the statement is incorrect* (after Mitrovic & Ohlsson 1999: 239).

One way to implement state constraints is as pairs of patterns that are matched to problem states (Mitrovic & Ohlsson 1999: 239). A two-step procedure can be used to determine whether a given problem state satisfies or violates the members of a set of constraints. In the first step, the relevance patterns of all constraints in the set are compared to the problem state in order to find the subset of constraints that are relevant in this state. The second step tests for each of the relevant constraints if its satisfaction pattern matches the problem state. If that is the case, the problem state satisfies the matching constraint; otherwise, it violates the constraint (pp. 239f.).

In instructional systems, constraints are used to both model the domain and to evaluate the solutions produced by the learner. The constraints represent syntactic or semantic properties of the domain being modeled (Ohlsson & Mitrovic 2006: 9). The current problem state has to be represented in a form that can be matched to the constraints (p. 9f.). While changes to the problem state have to be recorded, the actions or processes that are responsible for these changes do not (p. 10). The consistency of the problem state with the set of constraints is checked by comparing the state to the constraints as described above and noting any constraint violations (Ohlsson & Mitrovic 1999: 239; Ohlsson & Mitrovic 2006: 10). Since correct solution paths cannot involve problem states that violate constraints of the domain (Mitrovic et al. 2003), constraint violations indicate student errors and point to underlying misconceptions or missing conceptions (i.e. incorrect or incomplete knowledge, cf. Section 12.2.3.2), which can be addressed by appropriate instructional actions (e.g. feedback messages) (Mitrovic et al. 2001; Mitrovic et al. 2006). If a constraint is satisfied, the system does not intervene. Each constraint defines a class of problem states whose members are instructionally equivalent, i.e. they activate the same instructional response which can be directly linked to the constraint (Mitrovic et al. 2006; Ohlsson & Mitrovic 2006: 12). Since the student's current state rather than his or her sequence of actions is of interest, no solution path is prescribed: as long as no constraint violation is detected, the student can perform any actions he or she chooses (Mitrovic et al. 2003; Mitrovic et al. 2006). The short-term student model keeps track of the constraints that have been satisfied and the constraints that have been violated by the learner[204] and provides the basis for feedback to the learner during problem solving (Ohlsson & Mitrovic 2006: 10). For modeling the learner's long-term knowledge, overlay (cf. Section 12.2.3.1) or probabilistic (cf. Section 12.2.3.4) techniques can be used (p. 10).

Constraint-based modeling has several advantages over other, more complex approaches to student modeling. First, the process of student modeling essentially consists of pattern matching and is therefore more efficient (Mitrovic et al. 2001). Second, constraint-based modeling systems can be built with less cost and effort (Mitrovic et al. 2003). Laborious empirical studies to inform the construction of bug libraries (cf. Section 12.2.3.2) are not necessary (Ohlsson & Mitrovic 2006: 17). Constraints are modular (i.e. there is no direct interaction between them), general (i.e. they embody correctness criteria that are assumed to apply everywhere in the domain), and do not require a hierarchical or other kind of organization (Ohlsson & Mitrovic 2006: 8f.). As a result, constraint sets have turned out to be easier to represent, implement, and debug than bug libraries or rule sets (Mitrovic & Ohlsson 1999: 253; Ohlsson & Mitrovic 2006: 8f.). Third, correct student solutions are recognized even if they use a different approach than the ideal solution. If the learner's solution does not violate any constraints, it is regarded as correct with respect to the constraint database (Ohlsson & Mitrovic 2006: 13). Problem-solving steps that do not provide instructionally relevant information do not trigger any constraints, which means that they are effectively ignored by the system (Mitrovic & Ohlsson 1999: 240). In fact, CBM ignores the learner's particular problem-solving strategy altogether (Mitrovic et al. 2001) and is therefore open to free exploration and creative solutions (Abdullah 2003: 23). Fourth, complete coverage is not a requirement for constraint sets as long as the most typical learner errors can be detected (Mitrovic & Ohlsson 1999: 240). Fifth, constraint-based modeling has been argued to avoid the overspecificity problem of student modeling (i.e. the problem that techniques for modeling the student's knowledge tend to expect more detail than can be provided (p. 238)) because problem-solving steps can be evaluated selectively (and those which are not instructionally informative can be ignored), and the system works with classes of instructionally equivalent solution paths that can be mapped to the same instructional response (p. 240).

However, constraint-based modeling also involves a number of problems (Ohlsson 1992; Mitrovic & Ohlsson 1999: 253). One concerns the feasibility of developing sets of appropriate constraints for qualitatively different domains, such as algebra, geography, and linguistics (p. 253). There is also the possibility that the constraint set created for a given domain has insufficient coverage and thus too many errors made by learners may escape the system's attention (p. 253). In general, if the relevant constraints cannot or have not been modeled, incorrect solutions may be regarded as correct, based on the default assumption that the student is right until he or she is shown to be wrong ("innocent until proven guilty"). This, in turn, may lead to unfounded confidence on the part of the learner that he or she has mastered a given skill (Mitrovic et al. 2003; Ohlsson & Mitrovic 2006: 13). Hence, a number of implementations of constraint-based modeling have used more restrictive techniques that compare the learner's solution to a stored ideal solution (Ohlsson & Mitrovic 2006: 13). A negative side to being

[204] The literature on constraint-based student models does not seem to be consistent with respect to the exact contents of the (short-term) student model. Some papers state that this model only consists of the violated constraints (e.g. Mitrovic et al. 2003), others say that it includes both satisfied and violated constraints (e.g. Ohlsson & Mitrovic 2006), and still others contain both statements (e.g. Mitrovic & Ohlsson 1999: 239 + 246).

flexible with respect to the student's problem-solving strategy is that unlike cognitive tutors (cf. Section 12.2.3.5), many constraint-based tutors are unable to provide strategic problem-solving advice because they cannot solve the problems themselves, although they can detect and give feedback on the absence of necessary elements in a solution if they have access to a pre-specified ideal solution (Mitrovic et al. 2003). The disadvantage of not having a library of common student bugs in CBM is that these bugs cannot be recognized and addressed by giving specific feedback (Mitrovic et al. 2003).

In sum, like other student modeling techniques, constraint based modeling is more suitable for some domains than for others, due to its characteristic combination of strengths and weaknesses (Mitrovic et al. 2003; Ohlsson & Mitrovic 2006: 18), which were outlined in the previous paragraphs. According to Ohlsson and Mitrovic, case-based modeling can demonstrate its strengths in two categories of domains. The first includes domains in which problem states have complex internal structures that become more and more detailed in the course of problem solving. In these domains, state constraints can be used to identify inconsistencies between parts of complex problem states that result from errors made by the learner (Ohlsson & Mitrovic 2006: 18). The second category benefits from CBM's ability to tolerate different ways of solving a problem. It includes domains in which correct solutions do not require a particular order of problem-solving steps and the learner's behavior has to be diagnosed based on the final result rather than his or her sequence of actions (p. 18). Furthermore, CBM may be considered if well-defined problem-solving strategies do not exist for a given domain or if modeling them would involve prohibitive time and effort (Mitrovic et al. 2003).

12.2.3.7 Collaborative Student Modeling

The approaches to student modeling reviewed so far use internal representations of the student whose contents is not accessible to the learner and which are maintained by the (intelligent tutoring) system alone. In contrast, *collaborative student modeling* opens the student model for inspection by the learner and makes him or her an active participant in its construction (Bull & Pain 1995; Beck et al. 1997; Bull 1998; Bull et al. 1999). In addition to task-related exchanges, the learner interacts with the system to negotiate or collaborate on the contents of an *open* or *inspectable student model*[205] (Kay 1997; Hartley & Mitrovic 2002; Bull et al. 2005; Bull et al. 2007a), contributing information explicitly that the system cannot reliably obtain on its own, such as the learner's current goal or his or her level of confidence in a particular response or action, and possibly questioning assessments derived by the system and suggesting corrections (Bull & Pain 1995; Beck et al. 1997; Hartley & Mitrovic 2002). The resulting model is a synthesis of the viewpoints of the system and the learner with respect to the latter's abilities, goals, performance, etc. (Beck et al. 1997). Student modeling thus becomes an interactive process in which the learner, rather than being the passive object of scrutiny and diagnosis on the part of the system, plays an active role (Self 1994b: 4). Apart from student-system

[205] An *open student model* makes its contents accessible to the learner and possibly to other interested parties, such as instructors, peers, and pedagogical software agents. Users can at least inspect such a model, but they may also be able to change it (Bull et al. 2007b). Open student modeling can be as simple as providing a visualization of the student model (e.g. a skill meter), but there are also more ambitious efforts in which learners actively participate in shaping the student model by engaging in negotiation or collaboration with the system (Hartley & Mitrovic 2002). Various ways to present and interact with open student models have been explored (Bull et al. 2005).

interaction, collaborative student modeling can also include contributions from human or artificial peers (cf. Section 12.3.10) and from human instructors (Bull et al. 1999). In fact, in educational settings where the learner interacts with a variety of human and/or non-human agents (instructors, other learners, computer-based tutors, etc.), the instructionally relevant knowledge about the learner is not exclusively in the possession of a single individual or system but is actually distributed among potentially many agents that interact with the learner (McCalla et al. 2000). If learner-related information is fragmented among multiple different sources, trying to maintain a single, monolithic global model for each learner in one place appears both infeasible and ineffective (McCalla et al. 2000). An alternative approach to collaborative student modeling dispenses with the traditional idea of a single global student model and views student modeling as

> *the process of assembling and summarizing fragmented learner information from potentially diverse sources. This information can be raw data, recorded by a web application, partially computed learner models inferred by an ITS, opinions about the learner recorded by a teacher or peers, or a history of learner actions. (McCalla et al. 2000).*

This 'active' learner modeling approach (McCalla et al. 2000) is characterized by a shift in focus from the *product* to the *process* of student modeling. The student model is conceptualized as a function by means of which a modeling agent computes just in time relevant information about a set of learners for a specific purpose using the computational resources available. The model computed is meaningful only in the context defined by the parameters of the function (agent, learners, purpose, and resources) (McCalla et al. 2000). Model *management* (rather than model building) becomes the major concern of the modeling agent, which involves dealing with the retrieval of fragmentary information, its integration, and the interpretation of the resulting model (McCalla et al. 2000).

The most common motivation behind collaborative approaches to student modeling is that by involving students and other parties in the construction of the student model, more accurate models of learners can be created (Beck et al. 1997; Bull 1998). A second goal is to encourage learners' "reflection [(about their learning and the domain)] through discussion and negotiation of the student model" (Bull & Pain 1995). Collaborative student models attempt to project a faithful image of the learner, rather than an approximation (Bull et al. 1999). If this image is in conflict with how learners see themselves, it is assumed that they are encouraged to reflect at a metacognitive level on their own knowledge and misconceptions (cf. Section 12.2.3.2) and to engage in a menu-based (Bull & Pain 1995) or natural language (Kerly et al. 2006) discussion with the system in which they might challenge some of the system's assessments or ask for an explanation of them (Bull & Pain 1995; Bull et al. 1999; Hartley & Mitrovic 2002). In response, the system can provide rationalizations for its ratings of the student and, in turn, challenge him or her with questions concerning the disputed pieces of knowledge. Correct responses provided by the learner may then lead to a reassessment of the learner's abilities (Bull & Pain 1995). A third goal of collaborative student modeling is to give the learner a sense of being in control of his or her own learning, something which traditional intelligent tutoring systems have largely failed to achieve (Beck et al. 1997). Because the learner's perspective as communicated to the system carries some weight in shaping how the system views his or her abilities, the learner comes to perceive that he or she can influence the behavior of the system to some extent (Beck et al. 1997).

Collaborative student modeling and open learner models are the topic of ongoing research (Bull et al. 2007a; Dimitrova et al. 2007). In general, by allowing learners to contribute to the construction of a student model, collaborative modeling becomes more complex than other approaches which do not request input from the learner. This is because the system has to distinguish between its own view of the learner and the viewpoints provided by the learner and by other parties, and has to combine the different sources (Beck et al. 1997). A fundamental problem of encouraging learner input into student models is that these contributions may not be accurate. Learners may be over- or underconfident and therefore overstate or understate their abilities, or they may lack the ability to precisely specify their level of knowledge about the subject matter (Beck et al. 1997). Collaborative student modeling techniques must be prepared to deal with such cases. When multiple external sources have to be integrated into the student model, it is necessary to determine their relative importance for modeling the learner in the current context, resolve inconsistencies and contradictions between them, and combine the different views into a model that reflects the learner's abilities more accurately than what the system can build on its own (Beck et al. 1997; McCalla et al. 2000).

With respect to the design of open student models, a variety of issues have to be addressed. Some examples of research questions are provided below, which have been adapted from lists compiled by Kay (1997) and Bull et al. (2007b):

- What is the most appropriate format for making the student model accessible to the learner?
- How freely should students be able to choose the presentation format of their model?
- Which aspects of the student model should learners (or others) be allowed to inspect or modify, and when?
- What is the right level of detail for modeling and presentation?
- What terminology and structures are comprehensible to the target group of learners?
- How can the complexity of realistic open student models be managed?
- What are effective strategies for using open student models in teaching and learning?

12.2.3.8 Case-Based Student Modeling

Case-based reasoning (CBR) solves new problems by referring to knowledge about previously encountered similar problems (cases) and their solutions (cf. Section 12.2.1). CBR techniques have recently been applied to student modeling in intelligent tutoring systems; the approach is called *case-based student modeling (CBSM)* (Shiri et al. 1998a; Shiri et al. 1998b; Shiri et al. 1998c). The basic CBSM method is to present a problem to the student which he or she tries to solve by (a) retrieving a stored case that he or she deems most similar to the problem; and (b) adapting the solution of that case (Shiri et al. 1998a). The learner's solution is evaluated by the system to build the *knowledge component* of the student model, which captures different aspects of the learner's knowledge (see below) (Shiri et al. 1998a).

Figure 67 shows the process of case-based student modeling in detail. The case-based student modeler (contained in the dark-grey box) is depicted as a component of an intelligent tutoring system. According to Shiri et al., the modeling cycle starts by choosing a problem for the learner to solve from a database of problems, based on input from the student model and a set of domain concepts to be taught. To solve this problem, the learner first selects what he or she thinks is the most similar case and then adapts the solution associated with that case (Shiri et al. 1998a). In the *accessible solution (AS) mode* (Shiri et al. 1998b), the learner has access to the set of cases and their solutions stored in the cases database and generates a solution by choosing one of these cases and modifying its solution. This mode is appropriate for novice learners (Shiri et al. 1998a). In contrast, in the *unaccessible solution (US) mode* (Shiri et al. 1998c), the learner cannot access the cases and their solutions stored in the system. Hence, he or she has to retrieve a relevant case from his or her memory and recall the solution steps for that case (Shiri et al. 1998c). This mode is for advanced learners who solve the current problem

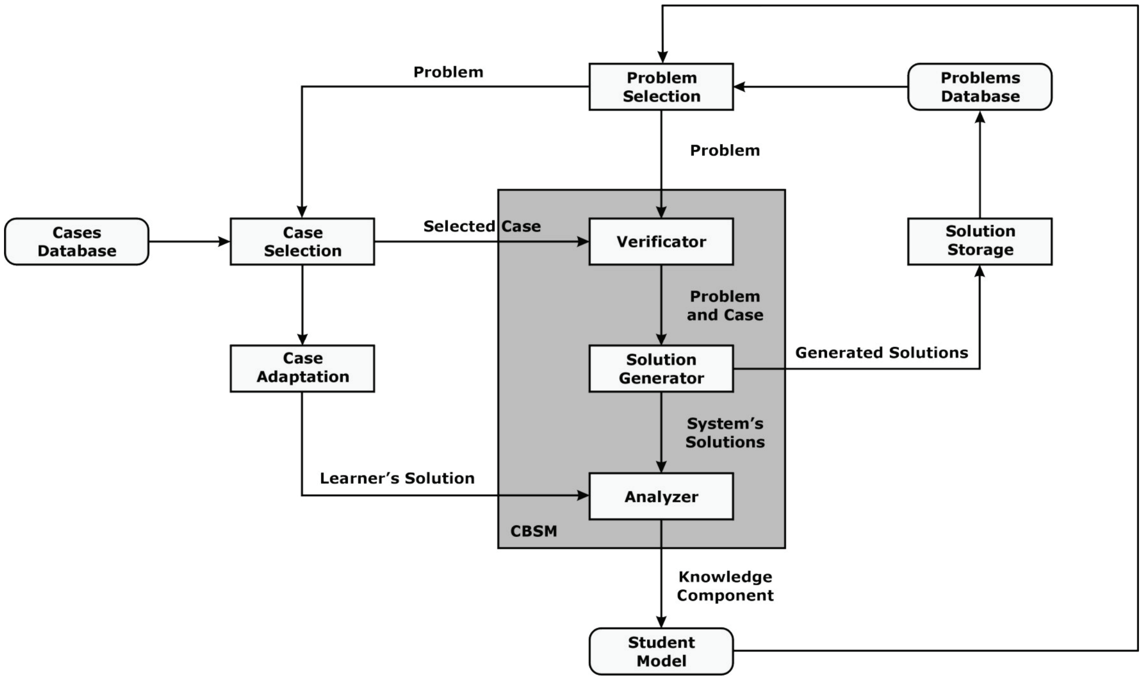

Figure 67. The process of case-based student modeling for the accessible solution (AS) mode. Adapted from Shiri et al. (1998a, Figure 3).

on their own by applying their knowledge of previous similar problems and their solutions (Shiri et al. 1998a; Shiri et al. 1998b).

The CBSM module for the accessible solution mode involves three processing steps, which are represented by three modules in Figure 67: the verificator, the solution generator, and the analyzer (Shiri et al. 1998a).[206] The first step is to make sure that the case picked by the learner is actually similar to the problem to be solved and therefore useful for creating a solution. If the case can be applied, it is used in the second step to generate a set of solutions to the problem; otherwise, appropriate feedback is provided. The representation of the problem in the problems database is updated with these solutions, so they can be referenced on future occasions and do not have to be re-computed. In the third step, the learner's solution is analyzed by comparing it to the solutions generated by the system in the second step, which are regarded as solutions that domain experts might produce using the same case (Shiri et al. 1998b). The results of this analysis are used to update the information in the knowledge component of the student model, which includes the level of the learner's knowledge (on a scale, cf. Section 12.2.3.1), his or her capabilities of retrieving and adapting relevant cases to solve problems, and both the problems solved by the learner (together with his or her solutions) and the problems that he or she did not solve. The latter are assumed to be useful indicators of where the learner may have difficulties (Shiri et al. 1998a; Shiri et al. 1998b).

Case-based student modeling provides a simple and computationally efficient approach to evaluate learners' problem solving, by looking at how they make use of knowledge of previous problems (cases) which are similar to the problem they are currently working on (Shiri et al.

[206] The architecture in Figure 67 shows the CBSM module for the accessible solution mode. In the unaccessible solution mode, the verificator is replaced by a discovery module, which uses heuristic search (cf. Section 12.1.3.1) to determine which case provided the basis for the learner's solution to the current problem. See Shiri et al. (1998c) for a detailed discussion.

1998a). In contrast to other student modeling techniques, the approach does not need complex inference algorithms (Shiri et al. 1998a). CBSM can be used to acquire different kinds of information about the learner's knowledge state, including knowledge level, problem-solving capabilities, and knowledge of solutions (Shiri et al. 1998a). Furthermore, it has been suggested that other components of the student model, such as a model of the learner's characteristics (e.g. his or her preferred method of reasoning), can also be built using case-based reasoning techniques (Shiri et al. 1998a; Shiri et al. 1998b). However, case-based student modeling still compares the learner's performance to the one of a domain expert (represented by the solution generator of the system) (Shiri et al. 1998b, cf. Section 12.2.3.2). Learners' creative use of cases to solve problems in novel ways which are not accounted for by the solutions generator will be rejected as incorrect. Furthermore, the approach seems to be restricted to the teaching of problem-solving skills (Shiri et al. 1998c). Finally, managing the size of the cases and problems databases can be a problem in complex domains.

González et al. (2006) described another way to use case-based reasoning in student modeling. In their approach, the student modeling component is implemented as a multi-agent system (cf. Chapter 14.1.5). The modeling process moves through the stages of the case-based reasoning cycle: retrieval, reuse, revision, and retention (cf. Section 12.2.1). These stages were described by González et al. and are summarized in the following. The learner to be modeled is regarded as a new case. The data about the student necessary to initialize the new case can be obtained via questionnaires, stereotypes, or tests (cf. Section 12.2.2). During the retrieval stage, the system tries to identify a similar, previously solved case (learner) by comparing the features of the new case against the database of stored cases. If the new case and a stored one are sufficiently alike (i.e. their degree of similarity exceeds a specified threshold), the diagnostic and problem-solving methods associated with the stored case are applied to (i.e. reused for) the new case. Revision (adaptation) of the selected case involves disposing of unsuitable items and removing inconsistencies. If no sufficiently similar stored case can be found, the case (learner) is treated as a completely new one and retained together with its specific information, including the results and misconceptions of the learner's problem solving, in the cases database (González et al. 2006). Based on a comparison of their case-based student modeling approach to student modeling with Bayesian networks (cf. Section 12.2.3.4), González et al. identified three advantages of case-based student modeling (González et al. 2006):

1. It facilitates handling, maintaining, and updating the student model.
2. Case-based reasoning techniques can be used for both instructional strategy selection and diagnosis of learner misconceptions.
3. The system can continuously model the learner's performance in terms of both qualitative and quantitative information.

12.3 The Instructor Model

Similar to human teachers, pedagogical agents are versatile, i.e. they can be designed to do many different things in order to promote learning. The *instructor model*[207] is the component of a pedagogical agent that embodies its instructional expertise in terms of a set of *instructional interactions* (i.e. joint instructional activities involving one or several agents and one or several

[207] The discussion of the instructor model in this section is based on and elaborates the one provided in Franke (2006: 271f.).

learners, in which both parties may participate in different ways and to different extents), which provide contextualized, qualified, personalized, and timely assistance, cooperation, instruction, motivation, and services for an individual learner or a group of learners (cf. Chapter 4). Given an instructional context and a set of instructional goals, the instructor model selects an appropriate instructional interaction and determines how to implement this interaction in order to achieve those goals in that context. The content, form, and timing of the agent's intervention are designed taking into account the learner's characteristics as well as the current state of the learning environment and the task, by interacting with both the domain model (cf. Section 12.1) and the student model (cf. Section 12.2).

Different instructional functions (roles) of pedagogical agents involve particular inventories of interactions, which include activities from among those to be discussed in the following subsections. The set of interactions extends the list provided by Clarebout et al. (2002: 270f.).

12.3.1 Supplanting

The agent completely takes over a task and carries it out for the learner, who observes the agent during its performance of the task (Clarebout et al. 2002: 270). The agent compensates for the learner's lack of capabilities and knowledge. Supplanting requires a complete model of the task and the environment (including the learner and other agents).

12.3.2 Scaffolding

In construction or maintenance, the term *scaffolding* refers to the act of setting up a temporary structure consisting of (interconnected) platforms arranged around a building, machine, or other construction exceeding a certain height, in order to elevate and support workers, equipment, and materials ([INT 149]). As summarized by Greenfield,

> *[t]he scaffold, as it is known in building construction, has five characteristics: It provides a support; it functions as a tool; it extends the range of a worker; it allows the worker to accomplish the task not otherwise possible; and it is used selectively to aid the worker where needed. (Greenfield 1984: 118).*

These ideas about the functions of physical scaffolding have been transferred to the domain of instruction. Scaffolding is a selectively supportive instructional strategy aiming to promote the extension of the learner's capabilities in the *zone of proximal development* (Vygotsky et al. 1978), which is the distance between what learners can achieve on their own and what they can achieve through interaction with more capable others (Vygotsky et al. 1978: 86; McLoughlin & Oliver 1998: 128; Kim 2003; [INT 14], cf. Chapter 2.2.3). The role of the teacher in instructional scaffolding is the following:

> *[T]he teacher's selective intervention provides a supportive tool for the learner, which extends his or her skills, thereby allowing the learner to successfully accomplish a task not otherwise possible. Put another way, the teacher structures an interaction by building on what he or she knows the learner can do. Scaffolding thus closes the gap between task requirement and the skill level of the learner. (Greenfield 1984: 118).*

Pedagogical agents which are more capable and knowledgeable with respect to the task at hand than the learner can offer this kind of scaffolding that enables the learner to cross the zone of proximal development and reach a higher level in his or her intellectual development

(McLoughlin & Oliver 1998: 128; Kim 2003). For example, agents may serve as cognitive tools (cf. Chapter 2.2.3) that enhance the cognitive abilities of learners by providing additional memory or processing capacities (Baylor 1999; Baylor 2001); share learning tasks as (more advanced) peers (Kim et al. 2006: 226f., cf. Section 12.3.10); or carry out those aspects of tasks that are beyond learners' (current) capabilities (Clarebout et al. 2002: 270).

12.3.3 Demonstrating

The agent first performs a particular task for the observing learner, and then the roles are reversed (Clarebout et al. 2002: 270). Embodied pedagogical agents can demonstrate tasks by carrying out actions involving the use of their body in the environment (cf. Chapter 4.1). Learners can view the demonstration from different perspectives, as often as they like, interrupt with questions, and ask to finish the task themselves (Rickel & Johnson 2000: 100). The agent can demonstrate a task under different conditions and with different parameters. Demonstrating requires that the agent be able to construct, revise, and execute plans that accomplish the task and to monitor the learner's performance of the task (Rickel & Johnson 1997; Rickel & Johnson 2000: 100).

12.3.4 Modeling

The agent accompanies its demonstration of a particular task with a running commentary on what it is doing, and why (Clarebout et al. 2002: 270, cf. Chapter 2.2.3). While performing the task, the agent interweaves its actions with explanations of the steps, strategies, and mental models involved (pp. 270f.). Embodied agents can deliver their explanations multimodally, using natural, unambiguous, and contextually appropriate combinations of speech, gestures, facial expressions, eye gaze, postures, and locomotion (e.g. Rickel & Johnson 1997; Rickel & Johnson 2000; Lester et al. 1999b; Lester et al. 2000, cf. Chapter 10.1). Where appropriate in the current task and situation, multimedia elements and presentations are included in the agent's explanation (cf. Chapter 4.2 and Section 12.3.12).

12.3.5 Coaching

The agent observes the learner while he or she is working on a task, imitating the agent's behaviors and (ideally) its reasoning, which were made explicit during the agent's previous demonstrations and modeling efforts (cf. Section 12.3.3 and Section 12.3.4). At certain points, the agent intervenes with additional modeling, corrective feedback, encouragement, reminders, etc., either at the learner's request or proactively (cf. Chapter 3.1.6) to seize opportunities for learning while the learner is performing the task. Coaching is thus part of a cognitive apprenticeship model of instruction that involves the agent and the learner in the roles of master teacher and apprentice, respectively. The agent models the skill to be learned for the student, coaches him or her during practice, and gradually fades its support as the student becomes more proficient (cf. Chapter 2.2.3).

The following aspects have to be considered when designing the interventions of a coaching agent (adapted from Kim & Baylor 2006: 588):

- *Type*. Informative, corrective, motivational, or scaffolding;
- *Scope*. Global, situation-specific, task-specific, or action-specific;
- *Timing*. Beforehand, immediate, delayed, just-in-time, or afterwards;
- *Learner characteristics*. Knowledge, preferences, affective state, personality, sociocultural background, etc. (cf. Section 12.2);
- *Learning objectives and contexts*.

Contextually appropriate and carefully timed interventions presuppose that the agent has a number of capabilities. These include the ability to keep track of both learners' actions and the changing state of the learning environment, to adapt its behavior and internal models based on the results of monitoring, to choose the right moment for interacting with the learner (cf. Chapter 7.15.8), and to devise and deliver instructional interventions in a way that considers the learner's cognitive and emotional state.

The need for proper timing of interventions by computing technology in general was stressed by Fogg (2003: 41–44). He formulated the *Principle of Suggestion* as follows:

> *A computing technology will have greater persuasive power if it offers suggestions at opportune moments. (Fogg 2003: 41).*

Following Fogg, pedagogical agents should be designed as *suggestion technologies* which provide assistance and information to change attitudes and/or behaviors "at the most opportune moment" (Fogg 2003: 41).[208] In general, finding or creating such a moment is a difficult task that involves the consideration of many factors, such as the current state of the environment and the learner's activities, the learner's disposition (mood, motivation, self-image, etc.), the agent-learner relationship, sociocultural conventions, and ethical considerations. As a result, general guidelines for timing agents' interventions right are difficult to obtain (p. 43). However, Fogg identified some moments in which people may be more open to persuasion from others (p. 43):

- When they are in a good mood;
- When they realize that their current world view is no longer in accord with reality;
- When they can act on a request or suggestion immediately;
- When they think that they owe someone something because they have received a favor, have made a mistake, or have denied a request.

12.3.6 Advising

The agent offers counseling or recommendations on what would be a prudent future course of action or way of conduct for the learner from the agent's perspective (OED 1998: 26). The advice is given from a position of perceived knowledge and/or authority (p. 26) in order to help the learner by addressing his or her uncertainty or lack of direction with respect to what to do or

[208] Opinions are divided on the question whether persuasion is "fundamentally good" or "inherently unethical" (Fogg 2003: 212). Fogg argued that the answer to this question depends on how people use persuasion on other people (p. 213). Persuasion can be abused to achieve ends that are deemed unethical in a given sociocultural context, but it can also be used for positive outcomes (pp. 212f.). The main responsibility lies with the designers of the persuasive technology or experience, but the recipients of attempts at persuasion are well advised to be vigilant and critical as well (Weiten et al. 2009: 192).

how to behave. Advice has a number of properties that distinguish it from other forms of directive behavior, such as commands or instructions (Lieberman 2001: 476):

- It can be less specific than commands.
- It may focus on particular aspects of the learner's behavior.
- It does not have to specify action but can suggest or modify it.
- It only has to solve part of the overall problem.
- The order in which advice is given or received is not required to be strictly sequential.
- It may critically appraise earlier behavior or guide future behavior of the learner.
- Good advice is context-sensitive (p. 481).

In general, the agent's advice is *personalized* (based on what the agent knows about the learner, cf. Section 12.2), *ongoing* (provided continuously and timely), and *non-coercive* (the learner can choose to follow, turn down, or disregard the advice) (Lieberman 2001: 478). Its form can be demonstrative, graphical, and/or verbal, depending on the agent's capabilities and the requirements of the situation (p. 478). The agent should not force its advice on the learner but grant him or her freedom to decide what to do with it (p. 477). The learner's decision to accept, reject, or ignore a particular piece of advice is influenced not only by its contents and way of delivery but also by the learner's overall perception of the advisor's expertise, authority, and trustworthiness (cf. Chapter 3.2.2).

In general, influential advisors are (perceived as) trustworthy and confident, and they are also known for their accuracy (Park & Catrambone 2007). A trustworthy advisor has proven both competent and reliable in situations in which listening to bad advice involves some risk (Park & Catrambone 2007). In particular, the advice given by trustworthy advisors is perceived by the advisees to be in their best interest rather than to the advantage of the advisor or a third party. Embodied interactive software agents, being computer programs and not people, might be perceived as offering more 'objective' advice and hence might be preferred in some advice-giving situations (Park & Catrambone 2007). However, if the agent advisor represents a company or some other institution, it may be more difficult for the user to trust this particular agent.

Confident advisors show that they believe strongly in the quality of their advice (Park & Catrambone 2007). Embodied interactive pedagogical software agents can signal confidence by modulating their voice (cf. Chapter 10.1.1), facial expression (cf. Chapter 10.1.3), and tone of language, and thus increase learners' acceptance of their recommendations (Park & Catrambone 2007). A confident agent both appears and sounds calm and assured rather than hesitating or nervous about a situation or a piece of advice.

Another dimension of the advisor-advisee relationship is rapport, or the emotional bond that may develop between the two over time (Park & Catrambone 2007, cf. Chapter 7.16). As the duration of relationships between humans and the agents advising them extends beyond single encounters, the emotional connection that people experience with agents might strengthen as well. Children have already been observed to form bonds with a pedagogical tutor agent, which had positive effects on their motivation (Massaro et al. 2000: 312; Oviatt & Adams 2000: 339f.; Cole et al. 2003: 1395).

12.3.7 Testing

The agent implements ways to assess (qualitatively or quantitatively) the learner's knowledge, abilities, aptitude, etc. with respect to different (aspects of) tasks or the subject domain in order to inform guidance and support activities as well as provide a basis for performance or progress reports (Clarebout et al. 2002: 271; [INT 150]). The agent's assessment can be *formative* (process-oriented, while the learner is engaged in some activity) and/or *summative* (outcome-oriented, after the learner has completed the activity) ([INT 151], cf. Chapter 15.3.2 and Chapter 15.3.3). These two types of assessment address the quality of the learner's approach and the quality of his or her accomplishment, respectively.

12.3.8 Motivating

Motivation can be defined as an internal condition of an individual that provides incentive and direction for behavior, in particular to achieve certain desirable states (e.g. satiation or success) or avoid undesirable ones (e.g. danger or failure) ([INT 152]; Ortony et al. 2005: 174). In learning processes, motivation plays a key role because it can (Ormrod 2003; [INT 152]):

- Focus learners on particular goals;
- Increase the effort and energy invested by learners in order to achieve those goals;
- Induce learners to start more activities and to see them through;
- Promote learners' cognitive processing;[209]
- Identify consequences that reinforce learner behavior;
- Positively influence learner performance.

Given the importance of motivation for learning and the critical amplifying/modifying role of emotions in motivation (Elliott et al. 1999; Cañamero 2003: 132f.; Rolls 2003: 17), agents that seek to motivate the learner require the ability to interpret, manage, and express emotions (Elliott et al. 1999), qualities that will be discussed in the next chapter. Pedagogical agents with emotional intelligence (cf. Chapter 13.4) can motivate learners by (Elliott et al. 1999):

- Showing that they care about the learner and his or her successes and setbacks (Picard et al. 2004: 261, cf. Section 12.2.2),[210] thus establishing themselves as "attachment figures" whose opinion of the learner matters to him or her;
- Tuning their behavior to the learner's emotional state;
- Giving praise, offering encouragement, and reducing frustration where appropriate;
- Transmitting their own (displayed) enthusiasm for the content to the learner;
- Promoting the learner's enjoyment of the learning experience through enjoyable interactions with a deep and appealing virtual persona;
- Winning the learner's affection and trust with the aim to develop a friendly and productive long-term social-emotional relationship (cf. Chapter 7.16).

[209] Schank and Neaman discussed three ways in which cognition may be affected by motivation. First, a learner may choose to participate or not in a learning activity, depending on his or her motivation. Second, motivation influences how much a participating learner will pay attention during the activity. The third impact of motivation arises from its close connection with the learner's goals. These goals, in turn, determine what is learned and how the learned information is organized in the learner's mind (Schank & Neaman 2001: 61–66).

[210] Lee et al. (2007a) found that a co-learner (peer) agent (cf. Section 12.3.10) that communicated caring for the learner through expression of empathy as well as supportive and encouraging verbal feedback inspired more trust and improved learning (according to recognition and recall memory tests).

12.3.9 Conversing

The agent involves the learner in a spoken or written natural language conversation (cf. Chapter 11.3) which it carefully orchestrates in order to achieve some instructional goal, such as:

- Making the learner articulate his or her knowledge about a topic;
- Identifying and remediating missing or incorrect knowledge of the learner;
- Guiding the learner through a problem-solving process or to new insights;
- Teaching the learner new information;
- Motivating the learner;
- Giving the learner the opportunity to practice social or communication skills.

One-to-one tutorial dialogue is commonly assumed to be more beneficial for learners than receiving group instruction or studying from a textbook (VanLehn et al. 2007: 3). VanLehn et al. conducted a number of experiments to test what they called the *interaction hypothesis*:

> *When one-on-one natural language tutoring, either by a human tutor or a computer tutor, is compared to a less interactive control condition that covers the same content, then the tutees will learn more than the non-tutees. (VanLehn et al. 2007: 5).*

However, the experiments indicated that the effectiveness of tutorial dialogue depended on whether the difficulty level of the instructional content matched the knowledge level of the learners or went beyond it (VanLehn et al. 2007: 3). Tutorial dialogue made no real difference (in comparison to low-interaction text-based control conditions) when novice or intermediate learners studied material designed for their level of knowledge (p. 50). However, for novice learners studying content for intermediate learners (i.e. content that was *above* their level of capability and thus more challenging), tutorial dialogue was found to be more effective than forms of instruction with less interactivity because it provided the necessary scaffolding (cf. Section 12.3.2) that enabled learners to cross the zone of proximal development (pp. 50f.). As summarized by VanLehn et al.:

> *In short, these data indicate that if students are given instructional content that is designed for students at their level of preparation, then tutoring has no advantage over studying text alone. However, if the content is in the students' [zone of proximal development], then tutoring has a big advantage over studying text alone. (VanLehn et al. 2007: 52).*

As discussed in Chapter 4.2 and in Chapter 11.3.3.2.1, conversations and other agent-learner interactions can be initiated and controlled by the agent, the learner, or both. System control may be too restrictive and user control too unpredictable for instructional agent-learner conversations; hence, it seems appropriate to use a mixed-initiative approach (Lester et al. 1999a), in which the agent and the learner share control over the dialogue, because it gives learners the opportunity to participate freely while the agent can still intervene when appropriate or necessary. Having said that, there are certain types of instructional conversations for which either system initiative or user initiative may be useful. For example, system initiative could be used for testing the learner (cf. Section 12.3.7) and user initiative for research that involves the learner querying the agent as a knowledge source (cf. Section 12.3.13).

The interventions of conversational pedagogical agents have to be designed and delivered with the learner's cognitive, emotional, and social development in mind. The *content* of an agent's contribution is determined based on cognitive considerations while social (cf. Chapter 7.16) and motivational (cf. Section 12.3.8) aspects need to be considered when deciding on the *manner* in which the agent should deliver its contribution (Mayer et al. 2006: 36). Manner of delivery matters because the content of a given contribution may not be neutral but may actually pose a threat to the learner's positive and/or negative face (p. 36, cf. Chapter 11.3.1). Hence, politeness is an important quality of pedagogical agents designed to be tactful and motivating because it can help them to reduce the potential impact of threats to the learner's face posed by certain kinds of messages (e.g. criticism, bad news, requests, suggestions, etc.) (pp. 36f.). Polite communication, in turn, may have positive effects on students' perception of the learning experience (e.g. the perceived difficulty of the instructional content) and of the agent as well as on their performance (Wang et al. 2005; Wang et al. 2008).

As a step toward developing principles for the design of polite conversational pedagogical software agents, Mayer et al. investigated whether learners can detect different politeness tones of tutorial suggestions and requests, which were expressed through different wordings of the statements. The suggestions and requests used in the study were of eight types (Mayer et al. 2006: 37) that were derived from transcripts of tutorial dialogues, based on the set of general politeness strategies defined by Brown and Levinson (Johnson et al. 2004d, cf. Chapter 11.3.1):

- *Bald on record*. The tutor explicitly tells the learner what to do.
- *Conventional indirectness*. The tutor attributes an explicit instruction to the computer.
- *Request*. The tutor expresses what he or she would like the learner to do.
- *Tutor goal*. The tutor states what he or she would do in the learner's situation.
- *Joint goal*. The tutor suggests an action to be performed together with the learner.
- *Question*. The tutor asks an indirect question that indicates a possible action.
- *Student goal*. The tutor offers a guarded suggestion of what the learner should want to do.
- *Socratic hint*. The tutor's guarded suggestion is embedded in a question.

The participants were asked to rate a set of tutorial statements (which they were to told to imagine as being delivered by an on-screen character (Mayer et al. 2006: 37)) on negative politeness (i.e. the degree to which the participants perceived the (imagined) agent making the statement as giving them the freedom to decide for themselves) and positive politeness (i.e. the degree to which participants thought that the agent imagined to be delivering the statement was cooperating with them) (p. 36). Commands, whether directly given or attributed to the machine, received the lowest ratings on both dimensions. The highest ratings with respect to negative politeness were obtained for guarded suggestions and Socratic hints while guarded suggestions and statements expressing a joint goal were rated most favorably regarding positive politeness (pp. 38f.). Furthermore, the rating patterns suggested that participants with less computer experience regarded politeness of suggestions or requests as more important than participants with a high level of computer experience (pp. 40f.). Overall, these findings have the following implications for the present work:

- Learners are sensitive to the tone of politeness exhibited by a tutorial statement (p. 41), which, interestingly, was successfully conveyed only by its wording in the study.
- Bald on record and conventional indirectness are both inappropriate strategies for polite conversational pedagogical agents (p. 42).
- Polite agents should rather use guarded suggestions and questions (and generally be more indirect in their statements) (pp. 38f.).

Along the same lines, Wang et al. (2005; 2008) reported the results of a number of Wizard-of-Oz experiments (cf. Chapter 15.3.2) that revealed what they referred to as the *politeness effect*: the performance, attitudes, and motivation of learners were positively influenced by a pedagogical tutor agent which used a model of socially intelligent tutorial dialogue based on

Brown and Levinson's theory of politeness to generate polite feedback that enhanced learners' social face and mitigated threats to their face, compared to an agent that gave direct feedback (Wang et al. 2008: 98f.). The effect was stronger for learners who preferred indirect feedback, were less experienced with computers, and did not have an engineering background (p. 108). These results indicate that beyond the mere presence of a pedagogical agent (the persona effect, cf. Chapter 5.4.5), the way it behaves toward and communicates with the learner can make a difference for the learning experience (p. 99). Therefore, Wang et al. argued that the social intelligence of pedagogical agents, in particular their use of politeness strategies, deserves more attention from agent developers (p. 98).

While it is important for a pedagogical agent to contribute to conversations with the learner, it is also critical for the agent to interweave talking with *listening* appropriately to the learner. This goes beyond trying to build a representation of the meaning of the learner's message and includes conversational techniques that influence not only the outcome of the conversation but also the learner-agent relationship, by showing that the agent cares about and respects the learner. Examples of these techniques include (Park & Catrambone 2007):

- *Nondefensive listening*. The agent signals attentiveness without interrupting the learner.
- *Active listening*. The agent demonstrates that it is really trying to understand the learner, for example by facing the learner, maintaining eye contact, giving backchannel feedback, asking questions, and summarizing the learner's contribution.
- *Disclosure*. The agent takes the initiative in sharing ideas, information, opinions, etc. with the learner, thus encouraging him or her to reciprocate.
- *Editing*. The agent converses politely with the learner (see above) and does not overreact when something goes wrong (e.g. when the learner's response is incorrect).

Humans tutoring other humans is a long-established and effective method of instruction. Not surprisingly, developers of conversational computer-based tutors have studied human-to-human tutoring to obtain ideas for the design of human-computer tutorial dialogues (e.g. Graesser et al. 1995). Person and Graesser (2003b) compiled a list of fourteen characteristics of (or 'facts' about) effective human tutoring, which they argued to be instructive for the design of computer-based tutors as well:

1. Human tutors mostly do not implement ideal tutoring strategies based on well-established theoretical approaches like case-based reasoning (cf. Section 12.2.1), Socratic tutoring, and Vygotsky's theory of development (cf. Chapter 2.2.3).
2. Tutors use a limited inventory of dialogue moves that refer to an immediately preceding contribution from the learner. The set of moves comprises pumps for more information, hints that become more and more specific, elaborations on learners' contributions, and splicing in the correct solution when students' answers are incorrect or incomplete.
3. Tutors provide feedback to student errors immediately rather than delay it.
4. Tutors' feedback to student errors is indiscriminately positive, negative, or neutral. Still, they avoid giving strong negative feedback in order not to discourage the learner.
5. Tutors generally do not remediate or give specific feedback to errors made by learners.
6. Tutors refrain from making both positive and negative remarks about learners' ability.
7. The number of questions asked by both tutors and learners in one-to-one tutoring settings is significantly higher than in classroom situations.
8. Most questions in a tutoring session serve to establish a common ground between the tutor and the student. Students typically ask common ground questions to make sure that their

own knowledge is valid, whereas tutors generally use such questions to assess the learner's knowledge.
9. Good students do not ask more questions overall, but they do ask a larger number of deep-reasoning questions.
10. To determine whether learners actually comprehend the instructional material, it is better to evaluate the quality of their answers than to ask them directly if they understand.
11. Tutors generate a large number of examples in the course of a tutoring session.
12. Tutor-generated examples are much more often used to explain difficult aspects of the subject matter and to test the learner's knowledge than to address student errors.
13. Most of the time, tutors present the learner with concrete and difficult examples that involve specific referents and challenge the learner.
14. Tutors primarily use pre-constructed (canned) examples (which they made up themselves or gleaned from textbooks or other sources). Spontaneously generated examples are much less frequent.

Although human tutors are often not trained in expert tutoring strategies and draw on a small and comparatively unsophisticated inventory of instructional interventions, they are surprisingly effective in their job, and students benefit from their interactions with them (Bloom 1984; Graesser et al. 1995; Person et al. 2001). While considering the above facts about human tutoring can simplify the design of computer-based tutors, and the resulting systems may still be effective, the facts also indicate opportunities for improvement, where computer-based tutors can be equipped with more sophisticated tutoring strategies, such as discriminating feedback, Socratic tutoring, and modeling-scaffolding-fading techniques, that enable them to perform better than the average untrained human tutor (Person & Graesser 2003b).

Boyer et al. conducted a corpus study to investigate how different characteristics of learners, including their performance level, self-efficacy, and gender, influence the structure of task-oriented tutorial dialogue, with the goal of obtaining insights that are useful for the design of tutorial dialogue systems (Boyer et al. 2007). They collected and analyzed a corpus of tutorial dialogues occurring during problem-solving activities in the domain of introductory computer science. Based on their results, Boyer et al. gave the following recommendations for tutorial dialogue strategies that accommodate learners with different performance and self-efficacy levels (Boyer et al. 2007):

- *Encourage reflection* of low-performing learners who are reluctant to request feedback.
- *Give adequate feedback* to low-performing or low-efficacy learners more frequently.
- *Confirm understanding* of low-performing learners after tutor hints or advice more often.
- *Give acknowledgements* to high-efficacy learners more often.
- *Maintain a comfortable conversational setting* for low-performing learners.

12.3.10 Interacting as Peers

The instructional interactions discussed in the previous sections are characteristic of individuals (people or agents) who play an instructor or tutor role and thus belong to a group with different characteristics than the learner. However, pedagogical agents can also play other roles in which they are perceived by the learner as similar to himself or herself in terms of attributes including age, status, and level of knowledge or ability, i.e. as *peers* belonging to the learner's own group. Learners have been argued and empirically demonstrated to obtain substantial benefits from interacting with their peers (Kim & Baylor 2006: 569), which include cognitive, procedural, and social knowledge and skills, as well as motivation (cf. Section 12.3.8) and self-confidence (Aïmeur & Frasson 1996; Kim 2003; Kim & Baylor 2006: 572). Students can learn both from and with their peers, by:

- *Cooperating with teammates*. The members of a team have to learn both their own roles and how to coordinate their behavior with other team members in order for the team to function smoothly as a unit (Rickel & Johnson 2000: 99).
- *Competing with rivals*. Rivals vie with the learner for attention, resources, success, etc. They challenge the learner in the task and set a level of performance which the learner has to surpass in order to win.
- *Learning with companions*. Companions compliment, encourage, empathize, and banter with the learner, but they can also provide valuable suggestions (Kim 2003).
- *Handling troublemakers*. A troublemaker is an unreliable learning companion whose actions or suggestions cannot (always) be trusted because they may be intentionally deceptive in order to challenge the learner's knowledge and self-confidence (Aïmeur & Frasson 1996; Frasson et al. 1997).
- *Imitating role models*. A role model is an individual who provides an example which the learner should emulate.
- *Learning from peer tutors*. A peer tutor is similar in status to the tutee but more advanced with respect to the subject of tutoring. Peer tutors are generally perceived as being closer to the level and status of the individuals being tutored; therefore, they appear less intimidating and more trustworthy as advice givers ([INT 153]).
- *Teaching peer tutees*. Learners tutor their peers, and, in the process of teaching them, learn about the subject matter themselves (Biswas et al. 2005: 364; Katzlberger 2005: 1). The learning-by-teaching process consists of making preparations for teaching, performing the teaching activities, and reflecting on feedback, observations, and results obtained during and after teaching (Katzlberger 2005: 4).

When human players are not available, pedagogical agents can take over any of these peer roles (teammate, rival, companion, troublemaker, role model, peer tutor, and peer tutee). Depending on their role, peer agents require different levels of expertise (above, below, or at the same level as the learner) (Dowling 2000; Kim 2003). The appearance (cf. Chapter 9) and behavior (cf. Chapter 10) of the peer agent persona must be designed in a way that facilitates its perception and acceptance by learners as a member of their peer group.

12.3.11 Sequencing

The agent dynamically creates and guides the learner through a sequence of instructional units and/or activities (cf. Chapter 2.3.2), based on its evolving model of the learner's knowledge, abilities, motivational state, and preferences (cf. Section 12.2). The next unit or activity selected by the agent may:

- Elaborate on the current topic;
- Cover a new topic of the curriculum;
- Introduce the learner to a related interesting topic outside the core curriculum;
- Address the learner's deficiencies with respect to the prerequisites of another unit or activity.

The selection of materials and activities depends on the set of topics to be covered, the prerequisites of the unit or activity, the priority of the unit or activity, the learner's history of successfully and unsuccessfully completed activities and units, and the learner's preferences.

12.3.12 Presenting

The agent displays information about a particular topic of instruction, using combinations of media and enhancing the multimedia presentation with appropriate spoken commentary and gestures (André et al. 1999b; Johnson 2003: 91–94; Rist et al. 2004, cf. Chapter 14.2.3).[211] In an interactive presentation, the agent may also answer learners' questions about the material being presented (Johnson 2003: 91). Settings for presentations differ with respect to the number of agents and learners involved as well as the degree of interaction among and between the human and artificial participants (Rist et al. 2004, cf. Figure 68):

- *Non-interactive presentation*. A single animated agent presents information to the learner. There is no interaction between agent and learner during the presentation (Rist et al. 2004: 378ff.).
- *Hyper-presentation/dialogue*. The learner can engage in some kind of dialogue with a single presentation agent, using menus or spoken or typed natural language (Rist et al. 2004: 378 + 381ff.).
- *Presentation teams*. The learner watches a presentation involving the on-screen interaction of several agent performers.[212] However, he or she cannot participate during the performance and remains a passive observer (Rist et al. 2004: 378 + 383–387).
- *Interactive performances*. Several agent performers interact among themselves and with the learner, who may both respond and take the initiative in the ongoing open multi-party conversation (Rist et al. 2004: 378 + 392ff.).
- *Multi-party multi-threaded conversations*. The conversational setting involves multiple on-screen performers and multiple learners that may have multiple conversations among themselves and with one another (Rist et al. 2004: 378f. + 392ff.).

Research in the field of multimedia learning supports the idea of having virtual presenters that use speech and animation, by providing empirical evidence for the modality principle of multimedia learning (cf. Table 1), according to which people learn better when presentations use both the auditory and the visual mode to convey information (Clark 2002; Johnson 2003: 91; Low & Sweller 2005).

[211] It should be emphasized here that although pedagogical agents can be used to present subject matters, they are not another type of (interactive) content, just as little as a human instructor can be reduced to what he or she teaches. Rather than being manipulated, pedagogical agents themselves manipulate (create, select, sequence, etc.) content to help learners while they are working with the content.

[212] Examples for the effectiveness of presentation teams come from the popular children's educational television show *Sesame Street* ([INT 154]), which often uses multi-character performances involving human and/or puppeteered performers as an entertaining method to teach its young audience a variety of authentic concepts.

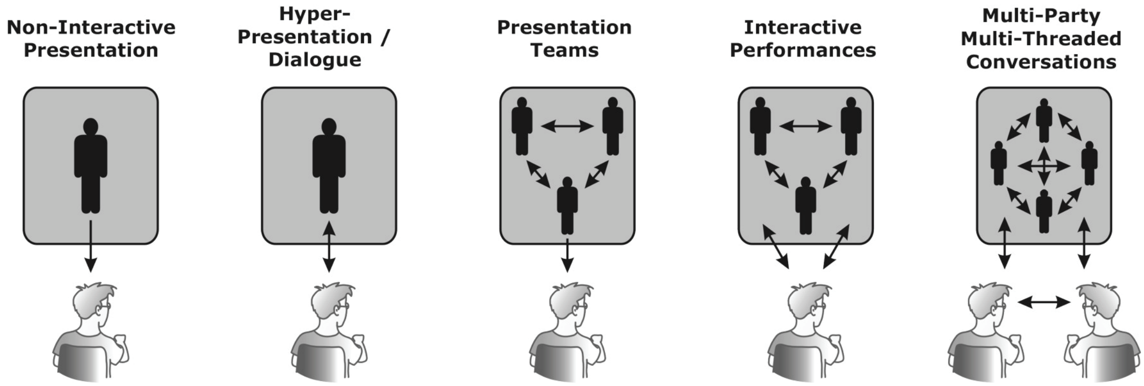

Figure 68. Character-based presentations in different conversational settings. Adapted from Rist et al. (2004: 379, Figure 1).

12.3.13 Knowing

The agent serves the learner as an expert or information source, providing relevant information about concepts, procedures, and resources on demand or unsolicitedly. The agent as *expert* shows mastery (extensive knowledge and/or above-average skill) in a particular subject area (Baylor & Kim 2005), which may include encyclopedic, instructional, problem-solving, and/or social expertise (see the discussion at the beginning of this chapter). Furthermore, experts generally appear confident, exhibit stability in performance, and maintain their composure despite unfavorable internal or external conditions (Baylor & Kim 2005).

The agent as *information source* aims to increase the learner's productivity by helping him or her to manage information and by protecting him or her from being flooded with information (Nwana 1996, cf. Chapter 3.2.3). This involves acquiring, collating, and manipulating different kinds of information, possibly in various formats and languages, from potentially distributed and heterogeneous sources. The information provided by the agent pertains to the learner's articulated or identified information needs and is delivered timely and in a format which the learner can understand and use.

12.3.14 Administrating

The agent manages organizational aspects of an individual's learning experience in the larger context of an e-learning course or curriculum. It usually works behind the scenes to create and maintain general conditions for learning that promote the learner's success. Examples of administrative activities include:

- Regulating access to instructional materials, activities, and resources;
- Reminding the learner of important events and deadlines;
- Keeping records of the learner's personal data, assignments, submissions, achievements, participation in course-related activities, etc.;
- Booking fee payments, selling additional printed materials, reminding payers in arrears, etc.;
- Helping with routine technical problems;
- Providing information about enrollment, course organization, requirements, navigation, how to use software, etc.

Pedagogical agents with administrative background tasks do not have to interact 'face-to-face' with learners and thus do not require an articulate embodiment or the ability to speak. Furthermore, they normally do not need teaching strategies or a detailed model of the student's knowledge (cf. Section 12.2). However, they must be equipped with knowledge and skills to deal with any routine issues related to the organization of the learning process.

12.4 Summary

Agents will be useful to, and sought by, users in a particular application context if they possess the relevant kind(s) of expertise (encyclopedic, instructional, problem-solving, and/or social) and successfully communicate their competence to the user. This chapter was concerned with the components making up the expertise of embodied interactive (pedagogical) software agents: the domain model (a repository of the knowledge which an agent has about its subject area), the student model (a user model built for educational purposes that captures the agent's evolving view of the learner and his or her development over time), and the instructor model (a representation of the agent's repertoire of instructional interactions).

Section 12.1 discussed the design of the domain model in terms of three knowledge-related activities: acquisition, representation, and use. The first is concerned with extracting knowledge for representation and use from sources including individuals, data, and experience. The second organizes, structures, and encodes knowledge in ways intended to facilitate acquisition and use. The third involves applying the acquired and represented knowledge to solve problems. Various methods for knowledge acquisition were reviewed in Section 12.1.1, including techniques for use by human knowledge engineers (cf. Section 12.1.1.1) and machine learning algorithms for the automatic acquisition of knowledge (cf. Section 12.1.1.2).

For the representation of the acquired knowledge in the agent's domain model, four major symbolic knowledge representation languages have been used, which were compared in Section 12.1.2: logic (cf. Section 12.1.2.1), semantic networks (cf. Section 12.1.2.2), frames and scripts (cf. Section 12.1.2.3), and rule-based systems (cf. Section 12.1.2.4). Section 12.1.2.5 explored the use of the Semantic Web approach to create external knowledge representations for agents in the form of semantic markup added as annotations to resources (objects, tasks, individuals, etc.) in different places in the agent's environment.

The use of the knowledge represented in the domain model for problem solving was discussed next in Section 12.1.3. Problem solving was defined as the ability of an agent to search for and find a sequence of actions that take it from a given initial state through a series of intermediate states to a desired goal state. Different search techniques were reviewed in Section 12.1.3.1. The following sections discussed several problem-solving approaches that are based on search: state space search (cf. Section 12.1.3.2), planning (cf. Section 12.1.3.3), and practical reasoning (cf. Section 12.1.3.4).

Student modeling was the topic of Section 12.2. First, Section 12.2.1 discussed aspects of user modeling by agents and other user-adaptive systems, including functions of user modeling (supporting system use and facilitating information acquisition), the acquisition and application of user models, their contents, direct and indirect strategies for obtaining information relevant to user modeling, and, finally, computational user modeling techniques (classification learning, collaborative filtering, decision-theoretic methods, plan recognition, and stereotypes). Next, in Section 12.2.2, the student model was defined as "an approximate, possibly partial, primarily qualitative representation" (Sison & Shimura 1998: 131f.) of various aspects of an individual learner, including his or her goals, plans, knowledge, beliefs, attitudes, emotions, personality, etc. A general account of the nature, contents, and functions of student models was provided in that section. Furthermore, a number of issues involved in the construction of student models (initialization, bandwidth, knowledge acquisition, knowledge type, student-expert difference, grain size, completeness and precision, level of analysis, time frame, inspectability, and sharability) were discussed in some detail. Section 12.2.3 reviewed different approaches to student modeling, which use scales and tallies (cf. Section 12.2.3.1); overlay, differential, and perturbation models (cf. Section 12.2.3.2); genetic graph models (cf. Section 12.2.3.3); probabilistic models (cf. Section 12.2.3.4); cognitive models (cf. Section 12.2.3.5); constraint-based models (cf. Section 12.2.3.6); collaborative models (cf. Section 12.2.3.7); and case-based models (cf. Section 12.2.3.8), respectively.

The third component of the expertise of a pedagogical agent was discussed in Section 12.3: the instructor model, which specifies the inventory of interactions that enable an agent playing a particular instructional role to provide one or several learners with contextualized, qualified, personalized, and timely assistance, cooperation, instruction, motivation, and services. Fourteen instructional interactions for pedagogical agents were discussed in this section: supplanting (cf. Section 12.3.1), scaffolding (cf. Section 12.3.2), demonstrating (cf. Section 12.3.3), modeling (cf. Section 12.3.4), coaching (cf. Section 12.3.5), advising (cf. Section 12.3.6), testing (cf. Section 12.3.7), motivating (cf. Section 12.3.8), conversing (cf. Section 12.3.9), interacting as peers (cf. Section 12.3.10), sequencing (cf. Section 12.3.11), presenting (cf. Section 12.3.12), knowing (cf. Section 12.3.13), and, finally, administrating (cf. Section 12.3.14).

13 Emotion and Personality

It may be more efficient to communicate through Internet and e-mail, but it is not possible to hug a computer. (Anonymous. Translated from German).

The question is not whether intelligent machines can have emotions, but whether machines can be intelligent without emotions. (Minsky 1985).

Humans are social beings whose behavior, thinking, and interactions with other individuals are considerably influenced by what they feel (their *emotions*) and who they are (their *personalities*). Not surprisingly, people also expect their interaction partners to exhibit consistent personalities and to handle emotions in a competent and natural way.

Interaction with computers has commonly been viewed as an exception: computers are the stereotypically impersonal, unemotional artifacts, which force users to suppress the affective and social parts of their nature while working with the efficient and 'rational' machine. However, even ordinary computers with simple text-based interfaces elicit social and emotional responses from users (cf. Chapter 6.5), a finding that has been replicated with embodied interactive software agents (e.g. Lester et al. 1997b; Nass et al. 2000a; Cole et al. 2003; Paiva et al. 2004).

Paraphrasing Ball (2003: 303), embodied interactive software agents will be *useful* if they add value to the user's experience (cf. Chapter 7.1); they will be *usable* (cf. Chapter 6.3) if they are effective and natural communicators. However, to be *comfortable* for users, they will have to present themselves as emotionally adept individuals with appealing and deep personalities, which invite the user to form a relationship with them. An agent that can address the user's emotional state and create (the illusion of) an interesting character requires a wide range of capabilities, including the following (Hudlicka 2003):

- Recognizing and interpreting the user's emotional state;
- Modeling the user's emotional state;
- Adapting to the user's emotional state;
- Generating appropriate emotional states within the agent;
- Communicating ('genuine' or pretended) emotions to the user;
- Portraying and revealing a rich and appealing agent personality and back-story.

These capabilities are discussed in this chapter. But first, it is useful to explore what 'emotion' (an intuitively clear but surprisingly complex phenomenon) means and involves in the context of the design of embodied interactive agents, by examining in some detail the two major views of emotion in humans, as a cognitive or as a physical phenomenon, in Section 13.1. After shedding some light on the nature of emotions (what they *are*), Section 13.2 identifies what they *do*, i.e. the functions of emotions in human cognitive processes and social interactions. Equipped with this understanding of emotion in humans, Section 13.3 turns to emotion modeling in computers, discussing communication-driven and simulation-based approaches. Section 13.4 then covers the interpretation and expression of emotions as essential component processes of agents' emotional intelligence. The remaining two sections of the chapter are concerned with aspects that define both human and artificial agents as distinct persons, including their characteristic psychological traits, or personality (cf. Section 13.5), and the story of their individual lives (cf. Section 13.6).

13.1 The Nature of Emotions

Emotion is a complex, multifaceted phenomenon which lacks a universally accepted definition (Kleinginna & Kleinginna 1981). For the purpose of the present discussion, emotions may be characterized as short-term changes in the internal state of an artificial or natural agent that provide "situationally and individually appropriate internal responses" (Ortony 2003: 189) to events or circumstances in the agent or in the environment which the agent regards as relevant to its needs, goals, or concerns. The plethora of theories on emotion can be largely classified in terms of the extent to which they emphasize the role of the body or the role of the mind (Picard 1997: 281f.; Picard 2000: 22):

- *Physical theories* are chiefly concerned with the bodily response (e.g. increased heart rate, hand perspiration, etc.) associated with an emotion and try to identify universal physiological cues corresponding to emotional states (e.g. furrowed eyebrows as a signal of anger). Physical aspects of emotion are discussed in Section 13.1.1 below.
- *Cognitive theories* focus on the mental aspects of emotion, in particular how emotions arise from mental *appraisals*[213] (evaluations) of the situation and how agents deal with emotions by changing or maintaining their relationship with the environment (*coping*). The cognitive component of emotion is covered in Section 13.1.2.

Either view of emotions – the physical just as the cognitive – is only a part of the overall picture. In fact, the generation of emotion and its experience involve the interaction of the body *and* the mind (Picard 2000: 22). Emotions in humans are characterized by neuroendocrine, physiological (visceral and muscular), behavioral, cognitive, communicative, social, and cultural components. Specific kinds of emotions or emotion episodes involve these aspects in different proportions (Cañamero 2003: 117).

13.1.1 Physical Aspects of Emotion

Research on the *physical* component of emotions focuses on the physiological responses (e.g. voice inflection (cf. Chapter 10.1.1), facial expression (cf. Chapter 10.1.3), posture (cf. Chapter 10.1.5), heart rate, pulse, temperature, blood pressure, and perspiration) that accompany emotions or quickly follow them (Picard 2000: 22). This approach goes back to the psychologist William James ([INT 155]), whose 1890 theory equates emotion with the mind's perception of physiological responses to a stimulus (James 1890). In other words, emotion is the experience of changes in the body, such as increased adrenaline level, heart rate, etc. James's famous example was that people do not fear a bear and then run from it but rather run from the bear and consequently fear it. The mind's perception of the bodily arousal is the emotion ([INT 155]).

While most theorists today acknowledge that emotions have a bodily component (although they may have different views of its nature), they reject James's view of equating emotion with bodily response (Picard 2000: 25). Still, the body is an important vehicle for *expressing* emotions (cf. Section 13.4.2). The term *sentic modulation* labels variations in motor output and

[213] For example, an agent's emotional state of anger may result from an appraisal such as the following: "I had to submit the online worksheet until 1 pm, but my roommate hogged the computer and thus prevented me from achieving this important goal. As a result, I am angry with him."

Table 19. Forms of sentic modulation. Adapted from Picard (2000: 27, Table 1.1).

Apparent to Others	Less Apparent to Others
• Facial expression • Voice inflection • Gestures, locomotion • Posture • Pupillary dilation	• Respiration • Heart rate, pulse • Temperature • Electrodermal response, perspiration • Muscle action potentials • Blood pressure

body readouts that serve as the primary means by which people (usually subconsciously) express their emotional state. Hence, physiological correlates of emotion created through sentic modulation are the main source from which software agents can obtain cues as to users' emotional states (Picard 2000: 25f.). Conversely, agents should use at least some of the channels of emotional communication available to people in order to convey their own affective state in human-agent interactions.

There are different forms of sentic modulation, which are more or less apparent to observers (cf. Table 19). Some forms can be perceived by watching or listening to another person while others are detectable through touch or use of measuring instruments (Picard 2000: 27). Most perceptible are the modulation of the facial expression (Ekman 1982, cf. Chapter 10.1.3) and the voice (Banse & Scherer 1996, cf. Chapter 10.1.1) (Picard 1995). For the former, it has been argued that there are basic emotions (cf. Chapter 7.15.7 and Chapter 10.1.3), each characterized by a distinctive set of facial muscle movement patterns.

13.1.2 Cognitive Aspects of Emotion

Apart from physical aspects, emotions also have a *cognitive* component. Not only are people often aware of their emotions, but emotions may also be generated based on prior cognitive assessment (Picard 1995). As a result, *appraisal theories* of emotion (Frijda 1986; Ortony et al. 1988; Lazarus 1991; Frijda 1993; Scherer 1999; Scherer et al. 2001; Ellsworth & Scherer 2003) have been developed, which emphasize the influence of cognitive variables and processes on emotion.

Appraisal theories view emotion as a mechanism for continuous, flexible adaptation to changes in circumstances that mediates between stimulus events and the agent's behavioral reactions (Petta 2003: 257). The emotional responses of a human or artificial agent are viewed as arising from a two-stage process (Gratch & Marsella 2003, cf. Figure 69):

- *Appraisal* evaluates the agent's relationship with the environment along several dimensions concerning the agent and generates an emotion.
- *Coping* deals with the generated emotion by suggesting problem-focused or emotion-focused strategies that may change the environment or the agent's beliefs or attention, respectively.

As the diagram in Figure 69 shows, there is a close interaction between appraisal, coping, and cognition, which develops over time. Cognitive processes build interpretations of the agent's current situation that provide the information necessary to appraise that situation. Coping uses the results of the appraisal process to select appropriate strategies for changing or preserving the agent-environment relationship and enlists cognitive processes to implement

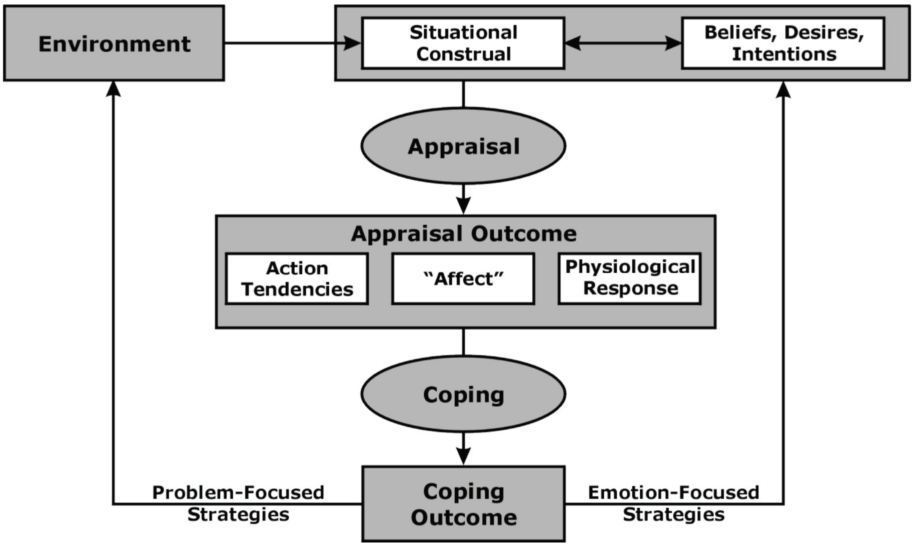

Figure 69. Typical components of an appraisal theory of emotion. Adapted from Gratch and Marsella (2004, Figure 1).

them. Eventually, coping strategies lead to new interpretations of the agent's relationship with the environment and possibly new appraisals (Gratch & Marsella 2004).

However, appraisal is by no means a *necessary* pre-condition for emotion (Picard 1995). In fact, neurological investigations have found evidence indicating that emotions can be activated without prior appraisal and can even temporarily seize control of the cognitive brain centers. For example, fear can make people jump out of the way of a rapidly approaching object before they can consciously perceive and analyze that a bus is about to hit them (Picard 2000: 32).

13.1.2.1 Appraisal

Appraisal is a process of continuous evaluation (Petta 2003: 257) by which agents assess their relationship with their physical and social environment, including their current state, the history of events leading to this state, and expectations of future developments (Gratch & Marsella 2005a). This assessment is formalized in terms of a set of abstract features for determining the affective state and guiding the behavior of an agent (Gratch & Marsella 2004). Appraisal is a reflexive rather than a deliberative process, but it both informs and is informed by cognition (Gratch & Marsella 2005a).

Cognitive processes involved in understanding and interacting with the environment (e.g. planning, explanation, perception, memory, linguistic processes, etc.) interpret external events in relation to an agent's internal dispositions (including its beliefs, desires, and intentions). Appraisal summarizes this interpretation by specifying the values of a set of *appraisal variables* (cf. Table 20), which describe the agent-environment relationship and are mapped to affective states (emotions and moods) of different type and intensity, including associated physiological responses and action tendencies (Gratch & Marsella 2004).

Appraisal variables can be regarded as criteria for judging the significance of events from the perspective of a particular agent. In other words, the significance of events arises only from their interpretation in the context of the beliefs, desires, and intentions of the agent, the agent's abilities, and past events (Gratch & Marsella 2004).

Table 20. Appraisal variables. Adapted from Gratch and Marsella (2004, Table 1).

Relevance		Does the event require attention or adaptive reaction?
Desirability		Does the event facilitate or thwart what the agent wants?
Causal Attribution	**Agency**	Who was responsible for the event?
	Blame and Credit	Does the causing agent deserve blame or credit?
Likelihood		How likely was the event; how likely is an outcome?
Unexpectedness		Was the event predicted from past knowledge?
Urgency		Will delaying a response make matters worse?
Ego Involvement		To what extent does the event impact the agent's sense of self (social esteem, moral values, cherished beliefs, etc.)?
Coping Potential	**Controllability**	The extent to which the event can be influenced
	Changeability	The extent to which the event will change of its own accord
	Power	The power of a particular agent to directly or indirectly control the event
	Adaptability	Can the agent live with the consequences of the event?

Appraisal serves as a mediator between events and emotions in the emotion generation process and thus accounts for the ability of the same event to elicit different emotions in different individuals or in the same individual (at different times). It also helps to identify the conditions that give rise to different emotions, and to understand the differences among emotions (Petta 2003: 257).

13.1.2.2 Coping

Appraising the situation, while an important first step, is not enough for an agent to achieve successful adaptation. The second step is for emotion to make an impact on the agent's actions (Petta 2003: 259). Emotions are not tied to specific actions but lead to the generation of *action tendencies*, which are defined as

> *states of readiness to achieve or maintain a given kind of relationship with the environment. They can be conceived of as plans or programs to achieve such ends, which are put in a state of readiness. (Frijda 1986: 75).*

For example, aggression is a tendency for individuals to behave when they experience anger. Other theorists have argued that action tendencies are not plans but rather innate biological impulses that may or may not lead to action, especially in humans (Lazarus 1991). Instead, they emphasize the role of *coping* in determining an agent's response to the appraised significance of events in the past, present, or future (Gratch & Marsella 2005b). In comparison to action tendencies, "coping is a much more complex, deliberate, and often planful psychological process" (Lazarus 1991: 114) that involves "cognitive and behavioral efforts to manage specific external or internal demands (and conflicts between them) that are appraised as taxing or exceeding the resources of the person" (p. 112). The coping process may suggest different strategies that change or maintain the relationship between the agent and the environment. Table 21 lists various coping strategies, categorizing them into two broad classes (Gratch & Marsella 2004):

Table 21. Coping strategies. Adapted from Gratch and Marsella (2004, Table 2).

Problem-focused Coping	**Active coping** Take active steps to try to remove or circumvent the stressor.
	Planning Think about how to cope. Come up with action strategies.
	Seeking social support for instrumental reasons Seek advice, assistance, or information.
Emotion-focused Coping	**Suppression of competing activities** Put other projects aside or let them slide.
	Restraint coping Waiting till the appropriate opportunity. Holding back.
	Seeking emotional support for social reasons Get moral support, sympathy, or understanding.
	Positive reinterpretation and growth Look for silver lining. Try to grow as a person as a result.
	Acceptance Accept the stressor as real. Learn to live with it.
	Turning to religion Pray, put trust in God (assume God has a plan).
	Focus on and vent Can function to accommodate loss and move forward.
	Denial Deny the reality of the event.
	Behavioral disengagement Admit inability to deal with the event. Reduce effort.
	Mental disengagement Use other activities to take the mind off the problem, e.g. daydreaming, sleeping.
	Alcohol/drug disengagement Consume substances that affect the metabolism in a way which changes the individual's perception of the problem.

- *Problem-focused* coping selects strategies that aim to change the environment. For example, the agent may deal with guilt by trying to make up for its mistakes.
- *Emotion-focused* coping deals with emotion internally, by proposing strategies that change how the agent interprets the circumstances. An emotion-focused way for the agent to cope with guilt would be to blame someone else. A changed perspective, in turn, can induce the agent to reappraise the situation, which may lead to anger toward the individual given the blame (Gratch & Marsella 2005b).

13.1.3 Primary and Secondary Emotions

The distinction between non-cognitively and cognitively generated emotions has been made in terms of primary vs. secondary emotions (Damasio 1994; Picard 2000: 35). *Primary emotions* (e.g. fear, anger, surprise, sadness, and happiness) are basic, innate ('hard-wired') emotions that form the underlying mechanism of secondary emotions. They are automatic and involuntary, i.e. usually do not arise from cognitive processing.

In human-computer interaction, objects and events on a computer screen may be responsible for automatic and involuntary (primary) emotional responses in users. Several examples were discussed by Brave and Nass (2003). For one, people may startle and experience sudden fear when a pop-up window, an animation, or an embodied agent appears or moves unexpectedly on the screen, or when the interface produces loud or sharp noises. Images can be arousing if they occupy much of the user's field of view, give the impression of coming toward the user, or move in the user's peripheral vision. Finally, humans are innately attracted to or repulsed by certain images and sounds, including explicit displays of sexual acts or violence and screeching or crying noises (cf. Table 9).

Secondary emotions (e.g. sympathy, guilt, pride, jealousy, and admiration) are higher-order emotions that arise as an individual matures and discovers that certain categories of objects and situations are systematically linked to particular primary emotions he or she experiences (Damasio 1994: 134). For example, grief involves bodily responses co-occurring with a cognitive understanding of loss (Picard 2000: 35). Secondary emotions may be the result of conscious or unconscious cognitive processing ranging in complexity from simple object recognition to sophisticated reasoning about the consequences of an action or event (Brave & Nass 2003). The majority of emotions relevant for human-computer interaction design are secondary emotions, including frustration, pride, and satisfaction (Nass & Brave 2005: 75).

13.2 The Functions of Emotions

Emotion is a necessary element of all intelligent agents, whether natural or artificial, for engaging with a complex and unpredictable physical and social environment. Humans, in particular, are the most expressive, emotionally complex, and socially sophisticated agents. They are equipped with two conceptually distinct, seemingly opposite but still complementary information processing systems: *cognition*, which is concerned with interpreting, reflecting on, and remembering aspects of the environment, and *affect*, which continuously provides quick positive or negative assessments of entities, events, and situations in the world (Norman et al. 2003: 38, cf. Chapter 6.4). Both systems are closely interrelated and depend on each other. They also play important adaptive roles in cognitive and social behaviors and processing, including decision making, planning, learning, attention, communication, relationship building, memory, etc. (Lazarus 1991; Damasio 1994; Picard 1995; Picard 2000; Brave & Nass 2003).

In artificial intelligence, the focus has traditionally been on emulating the cognitive abilities of humans, whereas affect in general and emotion in particular have been either completely disregarded or dismissed as a distraction to rational thinking (Gratch & Marsella 2005a). However, findings from neuroscience and psychology indicate that emotions are critical for aspects of behavior that have been traditionally viewed as 'rational,' including action selection, decision making, learning, and planning, among others (Hudlicka 2003: 10). Furthermore, not only does emotion influence cognitive behavior in various ways, but displays of emotion also serve important interpersonal functions and communicate an agent's internal state. *Affective computing*, i.e. "computing that relates to, arises from, or deliberately influences emotions" (Picard 2000: 3), is a growing field of research in artificial intelligence which acknowledges the importance of emotion in human-computer interaction and attempts to tap the potential of the emotion system with the goal to build 'intelligent' machines that can better make sense of and respond to complex and changing environments that include people, other machines, and themselves. The role of emotions in human and artificial intelligence was examined in a recent review by Martínez-Miranda and Aldea (2005), which summarized both research showing the importance of emotions in human intelligent behavior and research that seeks to incorporate emotions into intelligent systems, including affective agents.

In humans, and by analogy in artificial agents, emotion has two kinds of functions: *cognitive* (intra-agent) functions affect mental information processing, whereas *social* (inter-agent) functions mediate social interaction (Gratch & Marsella 2005a). Both functions are discussed in the following.

13.2.1 Cognitive Functions of Emotions

Emotion has strong adaptive influences on a wide range of cognitive processes in an agent. It is an important indicator of salience, guiding attention toward the important things and away from distractions, thereby helping the agent to manage its priorities (Picard 2000). It can facilitate or hinder memory and recall, and it regulates how agents solve problems and perform tasks (cf. Chapter 6.4). Specific cognitive functions of emotion include the following (Gratch & Marsella 2005a):

- *Situation awareness*. Emotion guides an agent's perception and categorization of salient inputs from its environment. In particular, it is instrumental in the process of interpreting the relationship between external events and situations on the one hand and the agent's internal beliefs, desires, and intentions on the other. The agent can then quickly determine whether subsequent ambiguous inputs are consistent with these prior appraisals, although this interpretation may be incorrect.
- *Action selection*. Emotion is involved in choosing and preparing an agent for a particular course of action in the environment. For example, fear causes arousal of the mind and body, directs attention toward the aggressor, and readies the individual for an appropriate action (flight, attack, etc.).
- *Coping*. Agents develop persistent cognitive and behavioral patterns (coping strategies) to deal with their emotions (cf. Section 13.1). Some of these are problem-focused (e.g. planning how to address a problem that causes negative emotions) while others are emotion-focused (e.g. denial, make-believe, and positive reinterpretation). The latter may be 'irrational,' but they still have positive effects on stress, life expectancy, and social-emotional relationships.
- *Learning*. Processes of learning in humans interact with emotions like curiosity, frustration, and satisfaction (Kort et al. 2001a). Memory and recall, problem solving, and decision making are influenced by emotional state. Positive emotions can facilitate learning, but encountering and surpassing failure and frustration is also an important part of the learning process (Burleson & Picard 2004, see also below).

  Kort et al. (2001a; 2001b) proposed a model of the interaction between emotions and learning (cf. Figure 70), in which learners' emotions change as they move through four quadrants (I – IV) and up a spiral during the learning process. In the model, the horizontal Emotion Axis ranges from negative to positive affect (emotion), whereas the upward section of the vertical Learning Axis represents the construction of knowledge (constructive learning) and its downward section the discarding of misconceptions (unlearning). Learners are constructing knowledge and experiencing positive affect in Quadrant I. Discrepancies between the study material and their own mental information structures cause learners to move to Quadrant II, which is characterized by constructive learning and negative affect

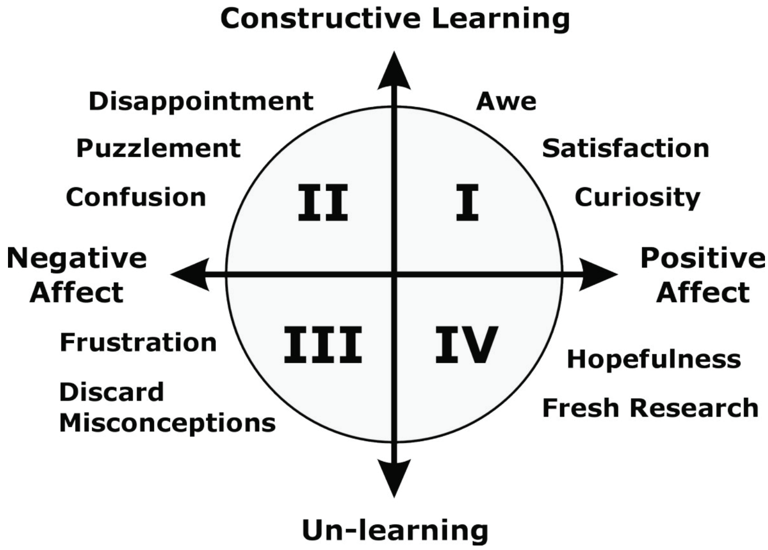

Figure 70. A cyclic model relating learning to emotions (Kort et al. 2001a, Figure 2; Kort et al. 2001b, Figure 1a).

(Craig et al. 2004: 242). Learners experience a *cognitive disequilibrium* (Piaget 1952, cf. Chapter 2.2.3) resulting from their encounter with stimuli or experiences that are not in line with their expectations, including contradictions, anomalies, obstacles, contrasts, surprises, perturbations, and equivalent options (Craig et al. 2004: 243). Cognitive disequilibrium is linked to the emotional states of confusion and possibly frustration, and it spurs cognitive processes of deliberation, inquiry, and search within the learner in order to restore *cognitive equilibrium*, which has the potential to result in deep learning. In Quadrant III, learners are eliminating their misconceptions about the study material (unlearning) while experiencing negative emotions, such as frustration. Quadrant IV is characterized by unlearning and positive affect. Having lost their misconceptions, learners now engage in new research, make new discoveries, and may even gain profound insights. They also recover hopefulness and positive attitude (Kort et al. 2001a). New ideas move learners back into the first (or second) quadrant, and a new learning cycle begins, which takes them further up the spiral (along the implicit Knowledge Axis).

Csikszentmihalyi (1975; 1990) argued that there is an ideal mental state of learning in which the student becomes fully absorbed in the learning activity, experiencing stimulation, complete involvement, and success ([INT 156]). This state, called the *zone of flow*, is characterized by clear goals, concentration and focus, a loss of the feeling of self-consciousness, a distorted sense of time, direct and immediate feedback, a balance between ability level and challenge, a sense of personal control over the situation or activity, the intrinsically rewarding nature of the activity, and full absorption of the individual in the activity ([INT 156]; Csikszentmihalyi 1975: 72). In the state of flow, materials and tasks have just the right level of difficulty for learners to forget time, fatigue, and boredom (Craig et al. 2004: 243).[214] Difficulty should not be avoided, but it "should be sought out, as a spur

[214] A proper level of challenge is also central to the user experience in computer games. Gee (2003) argued that the experience should be "pleasantly frustrating" for the user, i.e. challenging but not too much. Challenge can promote intrinsic motivation, not only in games but in all learning activities (Lepper & Henderlong 2000).

to delving more deeply into an interesting area. An educational system that tries to make everything easy and pleasurable will prevent much important learning from happening." (Kay 1991). In other words, *failure* is critical to learning (Schank & Neaman 2001: 38). Therefore, the possibility of failure should be explicitly designed into educational activities in order to create opportunities for learning. On the other hand, *fear* of failure constitutes a significant barrier to learning (Burleson 2005: 445). Schank and Neaman (2001) proposed several strategies for reducing this fear and keeping learners motivated:

- Minimize discouragement by decreasing the humiliation associated with failure.
- Help learners to understand that their failure will have minimal consequences.
- Motivate learners in ways that outweigh or distract from the perceived negative aspects of failure. For example, give them access to experts when they experience failure.

Pedagogical agents can be designed to be both empathetic and knowledgeable and thus have the potential to become caring experts (cf. Chapter 12.3.8 and Chapter 12.3.13) that contribute advice and encouragement in a non-humiliating fashion at the right time, minimizing the consequences of but not the opportunities for failure.

Learning processes are also greatly facilitated by learners' *intrinsic motivation*, i.e. when they deal with materials and activities for their own sake, without requiring external reinforcement (*extrinsic motivation*) (Sansone & Harackiewicz 2000). Indicators of higher levels of intrinsic interest include, on the one hand, more pleasure in the activity, more active involvement, and more task persistence, and, on the other hand, less boredom, less anxiety, and less anger (Craig et al. 2004: 243).

Craig et al. (2004) explored the role of learners' emotional states in learning with a pedagogical agent. They found significant relationships between learning and the states of boredom, flow, and confusion. Consistent with Csikszentmihalyi's model, learning gains were negatively correlated with boredom and positively with flow. Support for the idea that cognitive disequilibrium fosters deep learning was provided by the positive correlation between confusion and learning. In contrast, no evidence for a significant role of the basic emotions (anger, disgust, fear, happiness, sadness, surprise, and embarrassment, cf. Section 13.4.2) in learning was found (Kort et al. 2001a; Craig et al. 2004; D'Mello et al. 2006; Graesser et al. 2006; D'Mello et al. 2007a).

13.2.2 Social Functions of Emotions

Apart from mediating cognition, emotion also plays a number of important roles in the social capabilities of humans. In social interactions, emotional displays (cf. Section 13.4.2) carry information and regulate interpersonal behavior. While people deliberately portray emotions to manipulate others, communication of emotion is also instrumental in promoting cohesion and preventing misunderstandings in social groups (Gratch & Marsella 2005a). Emotional displays indicate when it is (not) appropriate or advisable to interact with others, and when it is, they indicate the kind of interaction expected (Ortony et al. 2005: 195). Furthermore, emotions are central to the development of long-term relationships between agents, restricting the self-interested behavior of an agent in favor of cooperative, altruistic behavior. Gratch and Marsella (2005a) discussed the following specific social functions of emotion:

- *Communication of mental state*. The emotional displays of others provide observers with a window into their beliefs (e.g. frowning may indicate disagreement), desires (e.g. joy shows the personal appreciation of an event or outcome), and intentions or action tendencies (e.g.

anger suggests aggression), although these displays may be based on communicative conventions rather than genuine emotional states. In addition, emotional displays convey information about dimensions relevant to the appraisal of the emotional significance of events, including valence, intensity, certainty, expectedness, and blameworthiness, among others.

- *Social manipulators*. Emotional displays can be used to manipulate other individuals into showing certain desired emotional responses. The targets may be unable to resist or may even be unaware of the manipulation. For example, manipulators may communicate anger to coerce action and enforce social norms; distress to obtain social support; guilt to reconcile with others; and joy or pity to signal that they feel happy or sympathetic for the addressee. It is also possible to influence the behavior of others by displaying emotions that give rise to particular emotional states in observers. Examples include *emotional contagion* ('catching' the emotions of others, similar to a disease)[215] and the *Pygmalion effect*.[216] Brave et al. (2005) investigated the psychological effects of two types of emotions in embodied agents: self-oriented emotion and other-oriented (empathic) emotion. They found that an agent that displayed empathic emotion was rated more positively by users in terms of likeability, trustworthiness, perceived caring, and felt support. Furthermore, agents that displayed empathic emotion were perceived as significantly more submissive than agents exhibiting self-oriented emotion or no emotion at all. In general, embodied interactive agents can be designed to enhance users' perception of being supported (by showing more other-oriented emotion) or of being independent (by showing more self-oriented emotion). Brave et al. suggested that designing the right level of agent support requires consideration of the user's personality and the application domain. For some applications, including education, they favored dynamic adaptation of agents' emotional behavior. For example, a pedagogical agent playing the role of coach (cf. Chapter 16.2.3) could provide more empathy and support for novice learners and make them feel more independent as they gain proficiency (Brave et al. 2005: 172f.).
- *Believability*. Convincing emotional displays of an embodied interactive agent can contribute to the believable, human-like image which it portrays to users and lead them to treat the agent in the same way as a human interaction partner. In particular, users interacting with a

---

215 People involuntarily mirror the facial expressions which they perceive in others. Apart from building connections and conveying empathy to their audience, this behavior can also cause people to experience the emotion that they mirror or 'fake' (Isbister 2006: 149). This is the basic idea of the *facial feedback hypothesis* (Darwin 1872; Strack et al. 1988; McIntosh 1996), which states that the experience of emotion can be influenced by facial muscle movements ([INT 157]). Likewise, recognition of a particular emotion in a voice can cause listeners to 'contract' that emotion as well (Nass & Brave 2005: 93f.). However, the complete emotion, including both arousal (high .. low) and valence (positive .. negative), is not always transferred to the audience. Listeners reliably catch the speaker's emotional arousal (excitement), which, in turn, intensifies whatever emotions they are experiencing at the moment, both positive and negative ones. In other words, a voice ringing with happiness can make a listener feel even angrier, more frustrated, more jealous, etc. Therefore, voices expressing strong emotions should be used with care in agent design, in order to avoid the danger of intensifying the wrong emotions in users (p. 94).

216 Verbal or non-verbal cues conveying certain (positive or negative) expectations of others about an individual can induce the individual to fulfill these expectations. This is known as the Pygmalion effect or the *self-fulfilling prophecy* (Gratch & Marsella 2005a; Draper 2006). The Pygmalion effect was originally observed in students who outperformed their peers simply because their teachers expected them to do so (Rosenthal & Jacobson 1992; Draper 2006).

believable agent are more polite, make socially desirable choices, are more nervous (Krämer et al. 2003), trust the agent more (Cowell & Stanney 2003), and feel greater empathy for the agent (Paiva et al. 2004).

13.3 Computational Models of Emotion

At the heart of an emotionally adept agent is a *computational model of emotion*,[217] which provides a symbolic description of human or human-like emotions, as well as their dynamics, at a certain level of granularity and in a format that facilitates recognition, interpretation, disambiguation, management, and synthesis of emotions in the context of the agent's application and/or sheds light on human emotion systems and processes (Elliott 2003: 237). Agents can use these models to (Hudlicka 2003: 6):

- Predict the user's emotional state or how he or she will most likely react to a (hypothetical) situation or event;
- Disambiguate incoming sensor data (e.g. does the user's increased heart rate mean that he or she is excited or afraid or angry?) or compensate for the lack of such data;
- Determine how to best respond to a recognized emotional state of the user.

Furthermore, computational frameworks of emotion can serve as a test bed for models of the mechanisms underlying emotional behavior and its influence on cognitive processing, which show how emotion is related to other mental systems like cognition and language (Gratch & Marsella 2004). Depending on whether computational models of emotion stress deliberately communicated emotion or the simulation of genuine emotion, they can be divided into communication-driven and simulation-based approaches (Gratch & Marsella 2005a):

- *Communication-driven* approaches use emotional displays as a means to achieve certain communicative or manipulative goals (cf. Section 13.3.1).
- *Simulation-based* approaches model aspects of emotion processes in order to simulate 'true' emotions in an agent (cf. Section 13.3.2).

13.3.1 Communication-Driven Emotion Modeling

In many implemented artificial agents, emotional behaviors are chosen on the basis of their desired impact on the user. As discussed by Gratch and Marsella (2005a), these agents do not 'have' emotions which arise from an internal computation of the agent's 'natural' emotional state given its current disposition and situation. Instead, emotional behaviors are selected which are useful at the current state of the agent-user interaction to achieve specific communicative, social, or task-related goals. Often, the agent's developers (rather than the agent itself) determine the utility of particular emotional behaviors and encode it into the agent's inventory of behaviors. Models for different agents and applications typically each have their own ad hoc implementation. Pedagogical agents have typically adopted a communication-driven approach, deliberately selecting emotions for their positive effects on student motivation and thus learning (Elliott et al. 1999).

[217] Enos (2003) provided an annotated bibliography of models of emotion that are useful in the context of computational applications.

Gratch and Marsella proposed that that the communicative function of emotional displays could be formally described in terms of *emotion acts* in analogy to speech acts (cf. Chapter 11.1.3.3). Emotion acts contribute information to the current state of the interaction and allow the same emotional display to be associated with several emotional acts and a given emotional act to be realized by different emotional displays (Gratch & Marsella 2005a).

Agents capable of portraying emotions will invite users to communicate back with emotional displays of their own. The ability to comprehend the user's emotional expressions provides agents with information about the user's cognitive or emotional state, the current state of the interaction, and the effectiveness of their communicative actions (Gratch & Marsella 2005a). Therefore, agents that communicate emotions also should be equipped with input channels and components for the detection and interpretation of users' emotional displays (cf. Section 13.4.1).

13.3.2 Simulation-Based Emotion Modeling

The focus of *simulation-based* emotion models is on the simulation of 'genuine' emotion in artificial agents, as opposed to the deliberate portrayal of useful emotional displays. Simulation-based approaches build computational models of cognitive processes from which emotional behaviors arise that provide adaptive responses to the agent's environment (Gratch & Marsella 2005b). Emotional expressions are not selected deliberately to achieve a communicative effect; instead, they originate from an internal representation of the agent's (simulated) emotional state and thus reveal something about its "inner life" (Gratch & Marsella 2005a).

The vast majority of current simulation-based approaches to emotion modeling draw on appraisal theories of emotion (cf. Section 13.1.2) as their conceptual basis, due to the popularity of these cognitive frameworks in psychological research today (Gratch & Marsella 2005b). The most prominent appraisal theory used in computational systems is the model by Ortony, Clore, and Collins (the OCC model). The Emotion and Adaptation (EMA) model extends the OCC model, integrating a plan-based model of appraisal with an elaborate model of coping. An alternative approach uses Bayesian networks to probabilistically model how emotion and personality are related to behaviors that express them. These three simulation-based models are discussed in the following sections.

13.3.2.1 The OCC Model

Andrew *O*rtony, Gerald *C*lore, and Allan *C*ollins' cognitive appraisal theory (Ortony et al. 1988) – commonly referred to as the *OCC model* – views emotions as the outcome of positive or negative responses of an agent to situations[218] involving events, agents, and objects, influenced by the agent's goals, standards, and attitudes (Picard 2000: 196). The model proposes a cognitive framework of emotions comprising specifications for 22 emotion types (cf. Figure 71). This structure groups emotions according to the cognitive conditions under which they are elicited (Ortony 2003: 193).

While it was originally intended for reasoning about emotions, the OCC model has become the default model for the generation of (cognitive) emotions in computers (Picard 2000: 195f.),

[218] The range of emotion-generating situations includes the experience of emotions itself. As a result, emotions can lead to further emotions or to the same emotion again and again (Picard 2000: 198).

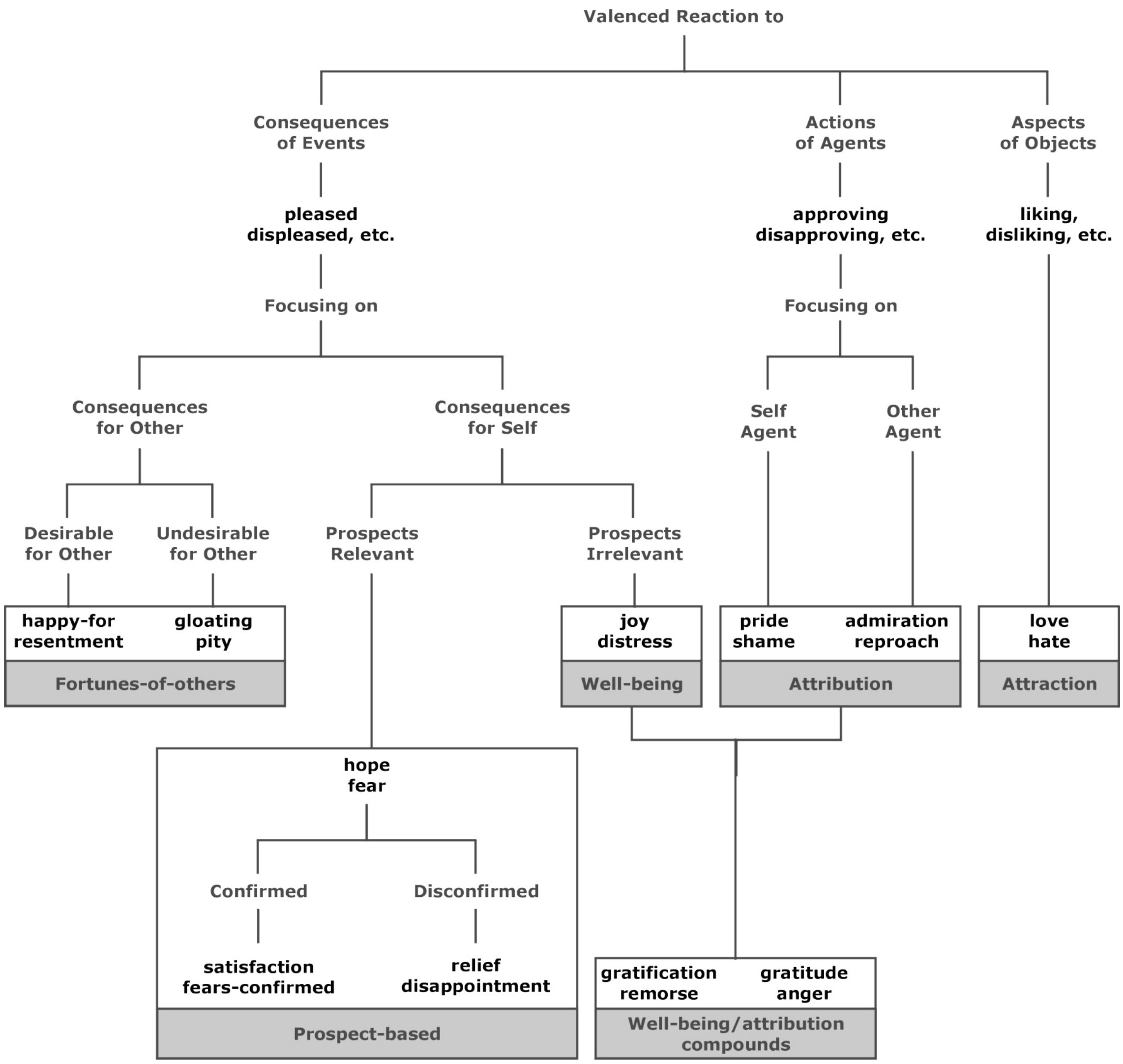

Figure 71. The cognitive structure of emotions as defined by the OCC model. Adapted from Picard (2000: 197, Figure 7.1). The 22 emotion types are given in black boldface print in the boxes along the bottom of the diagram.

which has found implementation in various computational systems and agents (e.g. Bates et al. 1992; André et al. 1999a; Elliott et al. 1999; Martinho et al. 1999). The OCC model provides a rule-based system for generating the different emotion types. The example below shows the rules for the generation of the emotional state of joy (Picard 2000: 196):

1. Set the potential P_j for generating a state of joy:
   $\texttt{IF } D(p,e,t) > 0 \texttt{ THEN set } P_j(p,e,t) = f_j(D(p,e,t),\ I_g(p,e,t))$
2. Set an intensity of joy I_j, given a threshold value T_j:
   $\texttt{IF } P_j(p,e,t) > T_j(p,t) \texttt{ THEN set } I_j(p,e,t) = P_j(p,e,t) - T_j(p,t)$
   $\texttt{ELSE set } I_j(p,e,t) = 0$
3. Map the resulting intensity to an emotion term from the joy family.

The first rule sets the potential P_j for synthesizing a state of joy. If the desirability D assigned by a person p to an event e at time t ($D(p,e,t)$) is greater than 0 (i.e. the consequences of event e are expected to be beneficial), then the potential of event e for generating a state of joy in person p at time t ($P_j(p,e,t)$) is computed by a function specific to the emotion of joy (f_j)

Table 22. Five specializations of generalized positive and negative affective reactions. Adapted from Ortony (2003: 194, Table 6.1). The undifferentiated positive or negative reaction is shown in the first section of the table. The remaining five entries are specializations (goal-based, standards-based, and taste-based).

Undifferentiated	
Positive Reactions	**Negative Reactions**
• Because something good happened (joy, happiness, etc.)	• Because something bad happened (distress, sadness, etc.)
Goal-based Specializations	
Positive Reactions	**Negative Reactions**
• About the possibility of something good happening (hope)	• About the possibility of something bad happening (fear, etc.)
• Because a feared bad thing didn't happen (relief)	• Because a hoped-for good thing didn't happen (disappointment)
Standards-based Specializations	
Positive Reactions	**Negative Reactions**
• About a self-initiated praiseworthy act (pride, gratification)	• About a self-initiated blameworthy act (remorse, self-anger, shame, etc.)
• About an other-initiated praiseworthy act (gratitude, admiration)	• About an other-initiated blameworthy act (anger, reproach, etc.)
Taste-based Specializations	
Positive Reactions	**Negative Reactions**
• Because one finds someone or something appealing or attractive (love, like, etc.)	• Because one finds someone or something unappealing or unattractive (hate, dislike, etc.)

based on its two parameters: the desirability `D(p,e,t)` and a combination of global variables affecting emotion intensity, such as expectedness, reality, and proximity (I_g`(p,e,t)`).

The second rule sets an intensity value I_j for the emotion of joy being generated. If the potential of event `e` for generating a state of joy in person `p` at time `t` (P_j`(p,e,t)`) is greater than a threshold value for activating the emotion of joy in person `p` at time `t` (T_j`(p,t)`), then the intensity of joy in person `p` given event `e` at time `t` (I_j`(p,e,t)`) is set to the difference between the potential and the threshold value; otherwise, the intensity is set to zero.

Finally, the resulting intensity value is used to find an appropriate emotion label from the joy family. For example, a moderate intensity value corresponds to 'pleased' while a very high intensity value translates to 'euphoric.'

Ortony later recognized the problem that the original system of emotion categories was too cumbersome for being implemented in affective artifacts (Ortony 2003: 193ff.) and proposed a refined scheme that reduces some of the categories in the original OCC model to five positive and five negative specializations of affective reactions (cf. Table 22). In addition, affective agents and systems need a coherent and stable value system for appraising their environment (p. 194). Such a system is presented in Figure 72. In this system, events, agents, and objects are appraised in terms of goals, norms/standards, and tastes/attitudes. The resulting appraisals give rise to emotions from different categories, including goal-based, compound, standards-based, and attitude-based emotions (p. 195).

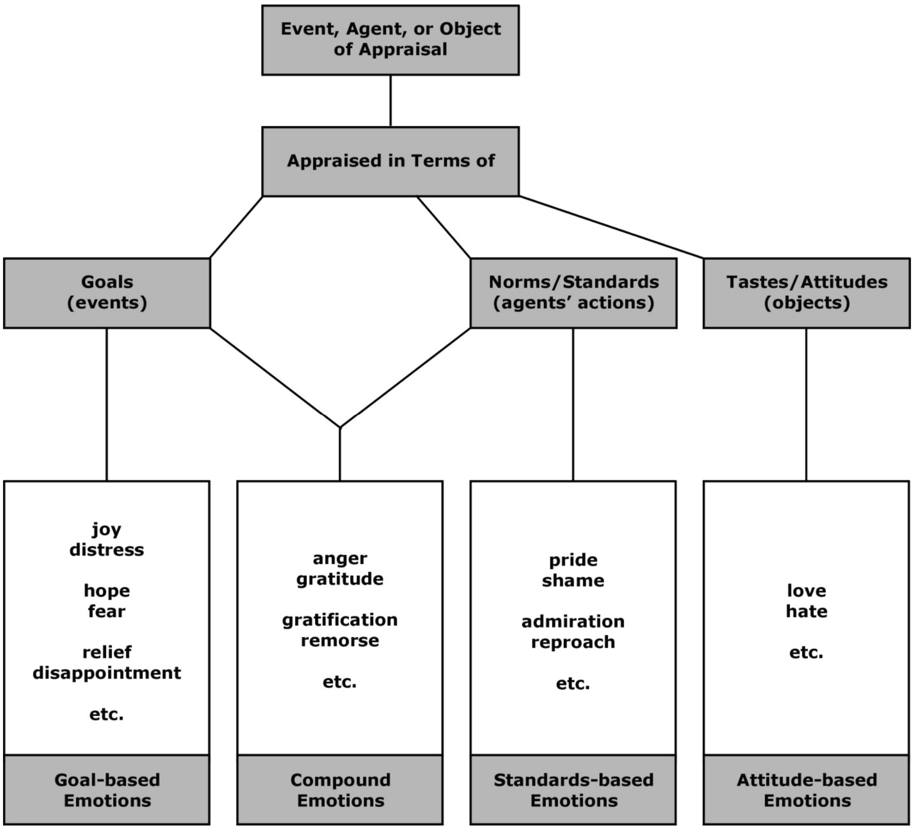

Figure 72. Appraisal and resulting emotions in the OCC model (Ortony 2003: 195, Figure 6.1).

13.3.2.2 The EMA Model

The *Emotion and Adaptation (EMA) model* has been designed as a general computational framework for modeling the mechanisms that give rise to different emotions in agents and determine how these emotions influence agents' cognitive and behavioral responses, with an emphasis on coping (Gratch & Marsella 2004; Gratch & Marsella 2005a; Gratch & Marsella 2005b). The model has been implemented in virtual humans (Gratch et al. 2002; Rickel et al. 2002; Marsella et al. 2004) to inform their decision making, perceptual attention, and non-verbal behavior (Gratch & Marsella 2005b).

Based on appraisal theories of emotion, EMA models the processes of appraisal and coping as a five-stage algorithm which manipulates a causal interpretation of the agent-environment relationship (cf. Figure 73). This interpretation is built by cognitive processes, characterized in terms of appraisal variables, and modified by coping responses (Gratch & Marsella 2005b). The EMA algorithm is summarized below (Gratch & Marsella 2004):

1. *Construct and maintain a causal interpretation of ongoing world events in terms of beliefs, desires, plans, and intentions.*
   The causal interpretation represents the agent's current mental state. It encodes past, present, and future actions and states; the likelihood and desirability of these actions and states; and causal relationships between them. Figure 74 shows an example, which is explained below.

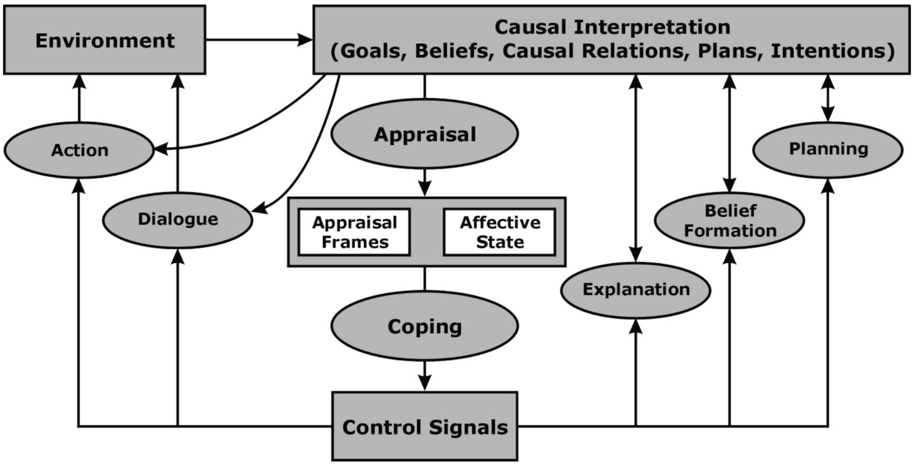

Figure 73. The Emotion and Adaptation (EMA) model (Gratch & Marsella 2004, Figure 2).

2. *Generate multiple appraisal frames that characterize features of the causal interpretation in terms of appraisal variables.*
   Each appraisal frame contains a set of appraisal variables whose values capture the assessment of a particular event from a certain perspective. New appraisals are formed and old ones are retracted depending on changes in the causal interpretation. Multiple events are appraised simultaneously and from different perspectives, including the agent's own perspective and the imagined perspectives of other agents.
3. *Map individual appraisal frames into individual instances of emotion.*
   This step uses mapping rules that take the configuration of variables in an appraisal frame and derive an emotion of a particular category (e.g. hope, joy, fear, distress, anger, or guilt) and intensity.
4. *Aggregate emotion instances into a current emotional state and overall mood.*
   Acting in the environment requires agents to perform a variety of operations (e.g. language processing, planning, perception, etc.) that read or update the causal interpretation. Access to elements of the causal interpretation brings their associated appraisal frames/emotion instances into focus as 'concerns' for the coping process (Gratch & Marsella 2003). The most-recently accessed emotion instances determine the agent's current emotional state. In contrast, the agent's mood results from aggregating all emotion instances connected to the current causal interpretation.
5. *Adopt a coping strategy in response to the current emotional state.*
   The selected coping strategy purports to maintain a desirable or reverse an undesirable emotional state currently in focus by identifying relevant items of the causal interpretation and assessing the potential for maintaining or changing them. Coping thus works with the same representation as, but in the opposite direction of, appraisal (Gratch & Marsella 2003).

The structure and elements of the causal interpretation, depicted in Figure 74, were described by Gratch and Marsella (2005b) as follows. The causal interpretation contains separate sections for past, present, and future (cf. Figure 74). It represents actions as rectangles, states as rounded rectangles, and causal relationships as arrows between actions and states. Each state indicates its utility for the agent (cf. Chapter 12.2.1), its current truth-value, its probability (indicating degree of belief for present states and likelihood of goal achievement for future states), and whether it constitutes a goal for the agent. Actions indicate who is responsible for their execution, the likelihood of their execution, and whether there are intentions to execute them. In

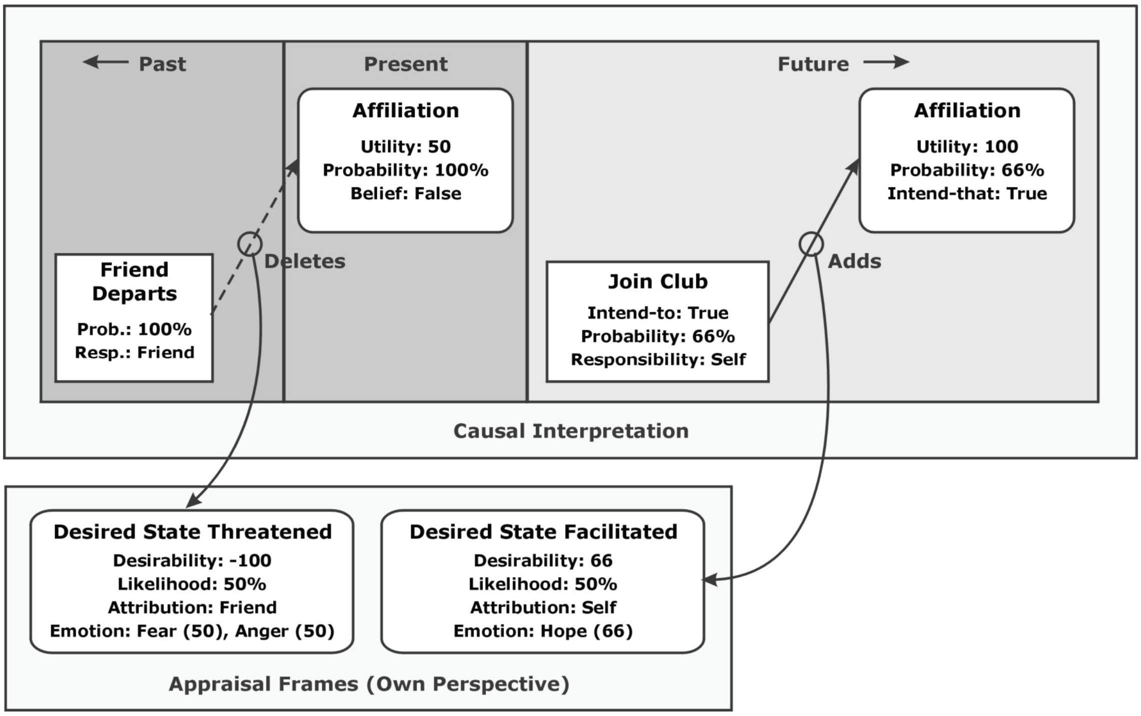

Figure 74. Causal interpretation and appraisal frames in the EMA model (Gratch & Marsella 2005b, Figure 3).

the causal interpretation shown in Figure 74, the agent has a current goal (affiliation) that was thwarted by the previous departure of a friend (the past action "Friend Departs" is responsible for the deletion of the Affiliation state). The agent may reachieve the Affiliation goal by performing the Join Club action.

Different actions have the potential to facilitate or hinder the achievement of desired states. Each case of facilitation or inhibition is assessed (from different points of view) in terms of an appraisal frame. Figure 74 shows appraisal frames encoding how the agent assesses the effects of two actions, namely the threat to the Affiliation state posed by the Friend Departs action and the future prospect of re-achieving the Affiliation state by means of the Join Club action (Gratch & Marsella 2005b).

13.3.2.3 Bayesian Networks

While the OCC model and the EMA model rely on complex, 'deep' processes of appraisal and coping to generate emotional responses, a simpler approach by Gene Ball and Jack Breese (Breese & Ball 1998; Ball & Breese 2000; Ball 2003) uses Bayesian networks (cf. Chapter 12.2.1) like the one in Figure 76 to model the inherently non-deterministic[219] relationship between emotion and personality and their behavioral expression (Ball & Breese 2000: 202ff.).

The approach by Ball and Breese reduces the complexity of both emotion and personality to two basic dimensions[220] in each case. To further simplify these continuous dimensions, they are

[219] That is, the relationship is hard or impossible to formalize in terms of explicit rules.

[220] Up to this point, emotions have been discussed as distinct *categories*. That is, they are either present or absent and can be (more or less accurately) labeled with a natural-language term. Alternatively,

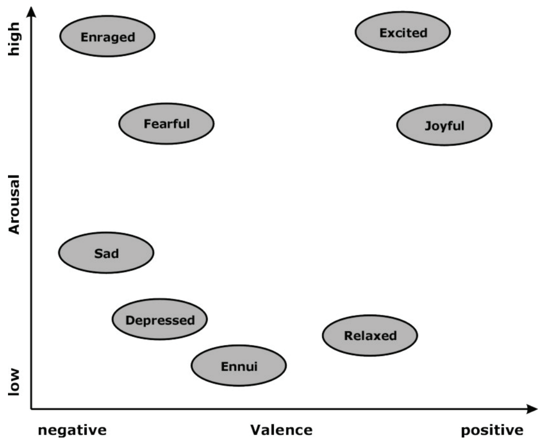

Figure 75a. Several labeled emotions located in the valence-arousal space (Ball & Breese 2000: 201, Figure 7.2).

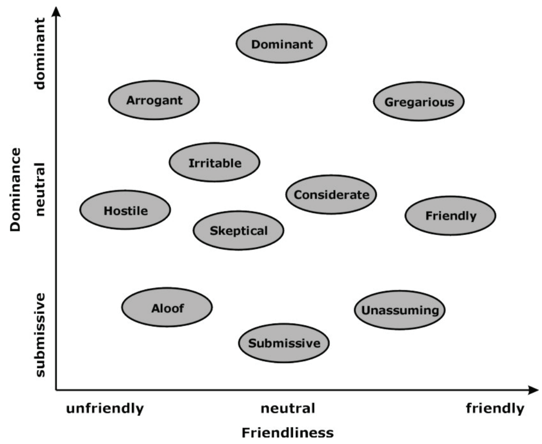

Figure 75b. The position of different labeled personality traits in the dominance-friendliness space (Ball & Breese 2000: 202, Figure 7.3).

characterized in terms of a small set of discrete values (Ball & Breese 2000: 200). Emotions (cf. Figure 75a) are specified with respect to the following dimensions (Ball 2003: 309):

- *Valence*. The positive-negative dimension (value: negative, neutral, or positive);
- *Arousal*. The degree of intensity (value: excited, neutral, or calm).

The representation of personality (cf. Figure 75b) consists of (p. 310):

- *Dominance*. How much an agent tends to control (or be controlled by) others (value: dominant, neutral, or submissive);
- *Friendliness*. The agent's disposition toward showing warmth and sympathy to others (value: friendly, neutral, or unfriendly).

As shown in Figure 76, a Bayesian network consists of nodes representing variables and directed arcs representing probabilistic causal interactions among the variables (cf. Chapter 12.2.1). Ball and Breese's network has internal, unobservable variables for arousal and valence (the dimensions of emotion) and for dominance and friendliness (the dimensions of personality), which are linked to observable variables representing aspects of verbal and non-verbal behavior (word selection, vocal expression, gesture, posture, and facial expression) that are influenced[221] by those hidden states with a certain probability, as specified by the parameters of the network (Ball & Breese 2000: 204f.; Ball 2003: 312ff.). This network has the advantage that it can be used for both the interpretation and the simulation of emotion. The operation of the Bayesian network as described by Ball (2003: 312) is summarized below.

emotions can be located as points (or regions) in a space defined by a small number of continuous *dimensions* (Picard 2000: 168ff.). As a result, the characterization of different emotions becomes a matter of defining the degree to which they vary along these dimensions (e.g. low .. high, positive .. negative). Important dimensions of emotion include valence and arousal; control is sometimes added as a third dimension (Nass & Brave 2005: 93).

[221] For example, the network in Figure 76 encodes the expectation that higher levels of arousal increase the pitch of speech.

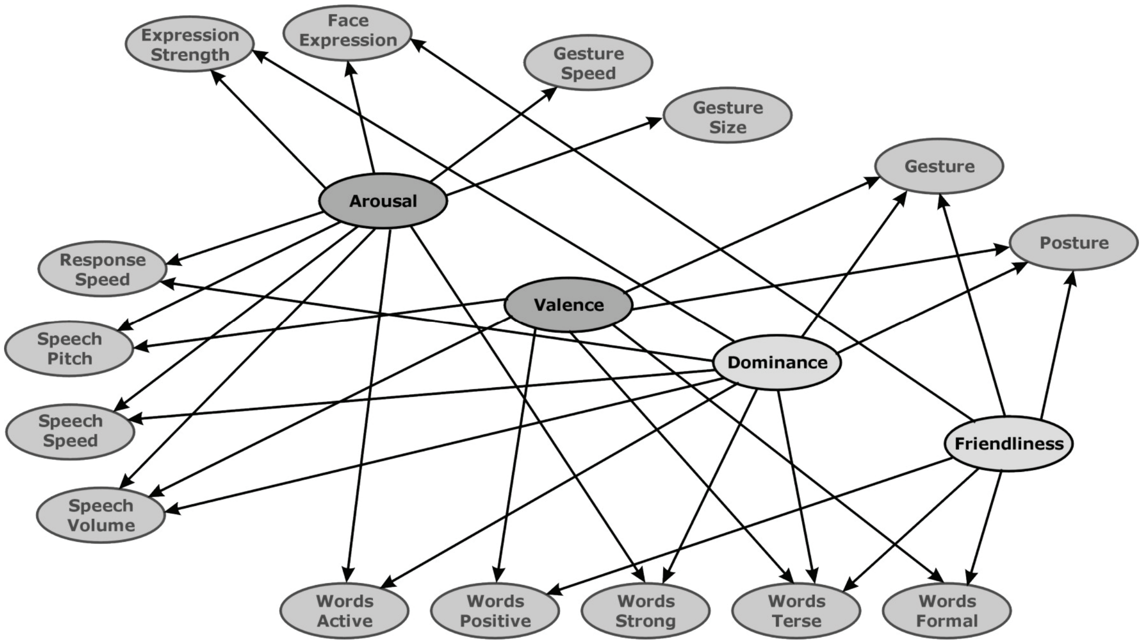

Figure 76. A Bayesian network modeling the relationships between emotion (represented by arousal and valence) and personality (represented by dominance and friendliness) on the one hand and their verbal and non-verbal expression on the other. Adapted from Ball and Breese (2000: 205, Figure 7.4) and Ball (2003: 313, Figure 11.2).

Interpreting the user's emotional state starts with observing the user's behavior and setting the values of (or instantiating) the corresponding behavioral leaf (or dependent) nodes in the network. Next, the network is evaluated, which produces estimates for the internal nodes corresponding to the components of the user's emotional state (arousal and valence). The result with the highest probability is interpreted as the user's perceived emotional state. Multiple results with similar probabilities render the diagnosis uncertain.

The *simulation* of behavior that is appropriate with respect to both emotion and personality operates in the reverse direction. The internal causal nodes of the network are instantiated with values representing the emotion and personality that the agent should portray. Evaluation of the network then gives probability distributions for the different available categories of verbal and non-verbal behavior. In other words, this evaluation estimates the likelihood of specific body postures, gestures, verbal messages, etc., depending on the specified personality type and emotional state. Random sampling of a probability distribution over time can be used to generate a sequence of behaviors that are consistent with the agent's personality and current emotional state but show a certain degree of variability (cf. Chapter 7.9).

13.4 Emotional Intelligence

Whether artificial agents (and computers in general) can *have* emotions, moods, etc. like humans do is a philosophical question that is probably unanswerable (Picard 2000; Picard 2003; Clark et al. 2004a; Evans 2004; Adolphs 2005), similar to the related issue of whether these artifacts can think (Turing 1950; Minsky 1982; Copeland 1993: Chapter 3; Horn et al. 1998). However, for human-agent interaction, the functional utility of emotion and of affect in general is of greater interest. *Emotional intelligence* (Picard 2000; Picard et al. 2001; Emmerling &

Goleman 2003) is the subset of social intelligence (Salovey & Mayer 1990) that includes the ability of agents to recognize, interpret, predict, express, and manage affective state (including emotions, moods, etc.) in themselves, the user, and other agents skillfully and for constructive purposes (Picard et al. 2001: 1175). Emotional intelligence is critical for the agent's effective cognitive and social functioning (cf. Section 13.2), together with the ability to make inferences from a model of personality (cf. Section 13.5) – its own and the personality of others. Most emotionally intelligent agent systems involve four components (Arnold 2003):

- The ability to interpret emotions (cf. Section 13.4.1);
- A computational model of emotions and their dynamics (cf. Section 13.3);
- Processes that guide behavior based on emotional state (coping, cf. Section 13.1.2.2);
- Channels for expressing emotions (cf. Section 13.4.2).

13.4.1 Emotion Interpretation

In social interactions, effective social functioning of artificial agents depends on their ability to *interpret* the emotions of the user, of other agents, and their own (simulated) emotions. External or internal emotional states are inferred based on observing the emotional displays of others and by way of reasoning about the current situation, which may include external events, actions of other agents, and aspects of objects in the environment (Picard 2000: 50). Alternatively (or additionally), it is possible to explicitly ask the user how they feel (user self-report) or to make assumptions based on information about user types or a typical user (Elliott et al. 1999, cf. Chapter 11.3.3.2.3). Characteristics of artificial agents with the ability to interpret emotions are listed below (Picard 2000: 55):

- *Input*. The agent receives visual, audio, and/or physiological signals from different input channels (cf. Chapter 11.2). The input signals may provide information about different forms of sentic modulation (cf. Section 13.1.1), including facial displays, voice inflection, gestures, posture, gait, respiration, electrodermal response, temperature, blood pressure, pulse, etc., depending on the array of sensors available.
- *Pattern Recognition*. The agent extracts features from the input signal and classifies them. For example, it analyzes an input video stream to extract features corresponding to different facial expressions (e.g. frowns vs. smiles), or it performs an analysis of voice quality, timing, and pitch contour to detect emotional arousal and specific emotions in speech (Picard 1997: 288).
- *Reasoning*. The agent infers the emotional state that most likely underlies the recognized emotional expression, using knowledge about emotion generation, the current situation, the user, and the sociocultural context.
- *Learning*. The agent improves its interpretation performance over time by acquiring knowledge about how emotions are generated and expressed from its interactions with particular individuals and contexts.
- *Bias*. The agent's simulated affective state (if applicable) influences its interpretation of ambiguous emotional displays. For example, an agent in a bad (or good) mood is more likely to interpret an ambiguous facial expression as negative (or positive) (Picard 2000: 54).
- *Output*. The agent provides a label or description for the identified expressions and their underlying emotions. Accuracy rates may be way below 100%, in the same range as humans, or perhaps even better if input channels beyond the human senses are used (Picard et al. 2001).

To be able to interpret audio-visual displays of emotion, agents have to be equipped with auditory and visual sensors for recognizing and tracking the user's facial expressions, gestures, and voice inflection (cf. Chapter 11.2).[222] In addition, information can be collected from other forms of sentic modulation (cf. Section 13.1.1) that are not accessible to humans. Examples include reading infrared temperature and measuring electrodermal response (Picard 2000: 50). Furthermore, current research is concerned with detecting the emotional state of learners who are engaged in a tutorial dialogue with a pedagogical tutor agent based on conversational cues extracted from the exchange, such as learner response time, number of characters and words per response, answer quality score, degree of tutor agent directness, and type of tutor feedback (D'Mello & Graesser 2007; D'Mello et al. 2008).

Facial displays are among the most perceptible indicators of underlying emotions (Cohn & Kanade 2007). Table 23 lists sets of observable facial cues for different (basic) emotions. However, speakers can also easily control their facial expressions to hide their emotions (Picard 1997: 288). The expression of emotional states is governed by *display rules*, i.e. cultural conventions that determine the appropriate manner, timing, addressees, and circumstances for different emotional displays (cf. Chapter 10.1.3). In addition, accurately mapping from observed facial expressions to underlying emotions can be difficult because a particular emotion may not generate overly expressive facial displays, and a given facial display may be ambiguous in the sense that it can be mapped to different emotional states, or the same (or a quite similar) display can be associated with different emotions. It is also possible that the facial image of the user that can be feasibly obtained given situational and technological limitations provides insufficient information for making an informed guess about the user's emotional state.

Apart from the face, a speaker's voice provides further important cues as to his or her current emotional state, including speech rate, pitch average, pitch range, intensity, voice quality, pitch changes, and articulation (Picard 2000: 179), whose different combinations and their corresponding emotions are discussed in detail in the following section on emotion expression. It is interesting to note that the *analysis* of emotions in speech has received much less attention in the speech processing community than the *synthesis* of emotional speech (Pantic & Rothkrantz 2003: 1379). The work that has been done has focused for the most part on the recognition of a subset of the basic emotions (cf. Chapter 7.15.7 and Chapter 10.1.3) from an input speech signal (Pantic & Rothkrantz 2003; Zeng et al. 2009). However, as discussed in Section 13.2.1, anger, disgust, fear, happiness, sadness, surprise, and embarrassment may be less important in learning than the non-basic emotions boredom, flow, confusion, and frustration. Ongoing efforts are concerned with detecting (and appropriately responding to) the latter emotions in an educational setting involving natural language dialogues between learners and a pedagogical tutor agent (Craig et al. 2004; D'Mello et al. 2006; D'Mello & Graesser 2007; D'Mello et al. 2007a; D'Mello et al. 2007b; McDaniel et al. 2007; D'Mello et al. 2008; Graesser et al. 2007; Graesser et al. 2008).

[222] See also Pantic and Rothkrantz (2003) and Zheng et al. (2009) for reviews of techniques for the audio- and vision-based automated recognition of human emotional states.

Table 23. Facial cues of different emotions. Adapted from Brave and Nass (2005: 141, Table 11.1.A).

Emotion	Observable Facial Cues
Surprise	Brows raised (curved and high)
	Skin below brow stretched
	Horizontal wrinkles across forehead
	Eyelids opened and more of the white of the eye is visible
	Jaw drops open without tension or stretching of the mouth
Fear	Brows raised and drawn together
	Forehead wrinkles drawn to the center
	Upper eyelid is raised and lower eyelid is drawn up
	Mouth is open
	Lips are slightly tense or stretched and drawn back
Disgust	Upper lip is raised
	Lower lip is raised and pushed up to upper lip or it is lowered
	Nose is wrinkled
	Cheeks are raised
	Lines below the lower lid, lid is pushed up but not tense
	Brows are lowered
Anger	Brows lowered and drawn together
	Vertical lines appear between brows
	Lower lid is tensed and may or may not be raised
	Upper lid is tense and may or may not be lowered due to brows' action
	Eyes have a hard stare and may have a bulging appearance
	Lips are either pressed firmly together with corners straight or down or open, tensed in a squarish shape
	Nostrils may be dilated (could occur in sadness too)
	Unambiguous only if registered in all three facial areas
Happiness	Corners of lips are drawn back and up
	Mouth may or may not be parted with teeth exposed or not
	A wrinkle runs down from the nose to the outer edge beyond lip corners
	Cheeks are raised
	Lower eyelid shows wrinkles below it, and may be raised but not tense
	Crow's-feet wrinkles go outward from the outer corners of the eyes
Sadness	Inner corners of eyebrows are drawn up
	Skin below the eyebrow is triangulated, with inner corner up
	Upper lid inner corner is raised
	Corners of the lips are drawn or lip is trembling

Recent research suggests that the consideration of (combined) cues from different input channels (cf. Chapter 11.2) may be the most effective way to detect different emotional states of users. For example, D'Mello et al. (2007a) reported that emotional states such as boredom and flow, which are not revealed by distinctive facial expressions, can be quite reliably detected using posture sensors, whereas detection of confusion and delight, which involve a high level of arousal, is best accomplished by monitoring the face. For the detection of frustration, D'Mello et al. suggested examining features of the conversation in the context of the human-agent dialogue. Overall, they found that high levels of detection accuracy can be achieved by using the appropriate emotion-sensor combinations. Furthermore, a combination of cues from multiple sensor channels can improve detection performance (D'Mello & Graesser 2007). Lee et al. (2007b) collected data about user actions, locations, temporal information, task structure, physiological response, and self-report affective states during interactions between learners and a task-oriented learning environment including interactive animated characters in order to induce different types of affect recognition models using machine learning techniques (cf. Chapter 12.1.1.2).

Once they have detected the presence of a particular emotional state in the user with sufficient confidence, agents require strategies that enable them to respond to this emotion in a way that takes into account the present context, the needs of the user, and cultural conventions and involves appropriate actions and emotional displays (cf. Section 13.4.2). To handle states of boredom, confusion, flow, and frustration, which play an important role in learning situations (cf. Section 13.2.1), the following strategies have been suggested for a pedagogical tutor agent (D'Mello et al. 2007a; Graesser et al. 2008):

- *Boredom*. Stimulate the learner's interest and cognitive arousal with challenging problems, options of choice, embedded games, etc.
- *Confusion*. Delay feedback and other support and intervene only if the learner fails to restore cognitive equilibrium (cf. Section 13.2.1) by himself or herself after a certain period of time.
- *Flow*. Keep a low profile (cf. Chapter 10.4) and challenge the learner within his or her zone of proximal development (e.g. with new or more difficult content and tasks).
- *Frustration*. Convey empathy (Baylor & Rosenberg-Kima 2006), correct misconceptions (cf. Chapter 12.2.3.2), and/or give hints that guide the learner's thinking toward more promising paths.

D'Mello et al. (2007a) noted the need for a pedagogical agent to guide the learner into a "virtuous cycle of flow and confusion," which keeps him or her engaged and promotes learning. At the same time, the agent has to make sure that the learner does not enter the "vicious cycle of boredom and frustration," which would be detrimental to both the learner's engagement and his or her success in the learning process.

13.4.2 Emotion Expression

In addition to interpreting and modeling emotion, affective embodied interactive agents require the ability to *express* emotions in a manner that is easily interpretable and convincing (believable) to the user[223] and observes the display rules of the target culture. This involves using at least some of the channels of emotional communication by which people convey their own affective state in human-agent interactions. The following characteristics of emotion expression in artificial agents can be identified (Picard 2000: 60):

- *Input*. Instructions about the emotion(s) to be expressed may be provided by the agent's own emotion processes, an external computer program, or a human being.

[223] People tend to ascribe emotions to others even if they do not display any visible cues. Hence, it is important for artificial agents to make their emotional expressions as clear as possible in order to reduce the risk that users misinterpret their behavior (Gratch & Marsella 2005b).

- *Intentional vs. spontaneous pathways*. Emotion expression may be the result of deliberation about whether or not the agent should communicate emotion (in response to the user's emotional state), and if so, how (*intentional path*). In contrast, the *spontaneous path* is available in an agent with simulated emotions, where the current simulated emotional state automatically affects selected outputs of the agent.
- *Feedback*. The emotional state influences emotion expression, but the expression may also feed back into the state.
- *Bias-exclusion*. The current emotional state is easiest to express for the agent, and it can inhibit the expression of other states.
- *Display rules*. Sociocultural conventions regulate the expression of different emotions (cf. Chapter 10.1.3).
- *Output*. Agents can portray emotional expressions through combinations of voice, facial expression, gesture, posture and locomotion displayed by an animated character as well as through appropriate music and background colors.[224] An agent has to know when to be more subtle and when to be bolder in its emotional expression.

Effective means for both humans and embodied interactive agents to express emotions include (combinations of) facial expression, voice inflection, gesture, and posture (cf. Chapter 10.1), with the face and voice providing the most perceptible cues (Picard 2000: 26f.; Russell et al. 2003).[225] The face is the primary means for visual communication of emotion (cf. Chapter 10.1.3). Most existing systems for either generating or recognizing facial expressions are based on the *Facial Action Coding System (FACS)* developed by Paul Ekman and Wallace Friesen to classify human facial expressions (Ekman & Friesen 1978; Ekman et al. 2002; Ekman & Rosenberg 2005; [INT 158], cf. Chapter 11.2). The FACS describes basic emotions (cf. Table 24) in terms of combinations of so-called *action units (AUs)*, i.e. primitive facial muscular movements which, in concert, generate different facial expressions (cf. Table 25) (Picard 2000: 175).

There has also been work on the expression of emotion in natural and machine-generated voices (e.g. Cahn 1990; Kappas et al. 1991; Murray & Arnott 1993; Johnson et al. 2002; Oudeyer 2003; Schröder 2004). In general, humans can detect emotions both in human and synthetic voices quite well, with recognition accuracies of 60 percent for human speech (Scherer 1981) and 50 percent for synthetic speech (Cahn 1990). Researchers have identified signatures of emotions in voice that consist of patterns of cues, including speech rate, pitch average, pitch range, intensity, voice quality, pitch changes, and articulation (Isbister 2006: 185f.). Table 26 summarizes cue patterns that signal different basic emotions in voices. A similar list of vocal cues for different moods and emotions is provided by Isbister (2006: 186):

---

[224] Juslin and Laukka found that "there are emotion-specific patterns of acoustic cues that can be used to communicate discrete emotions in both vocal and musical expressions of emotions" (Juslin & Laukka 2003: 799). The makers of films and computer games exploit the connection between music and emotion to intensify the viewer's or player's experience of the emotional reactions displayed by movie or game characters (Isbister 2006: 187f.).

[225] While visual behaviors for expressing emotions are already quite sophisticated, it is still difficult for current speech synthesizers to convey emotions properly (cf. Chapter 11.1.2.4). Furthermore, little research has addressed *affective language generation*, i.e. the task of communicating emotions through text and sentence structure or word choice (but see the annotated bibliography compiled by Piwek 2003).

Table 24. Basic emotions described in terms of action units. Adapted from Pelachaud (2003, Table 4).

Basic Emotion	Action Units
Anger	AU2 + AU4 + AU5 + AU10 + AU20 + AU24
Disgust	AU4 + AU9 + AU10 + AU17
Embarrassment	AU12 + AU24 + AU51 + AU54 + AU64
Fear	AU1 + AU2 + AU4 + AU5 + AU7 + AU15 + AU20 + AU25
Happiness	AU6 + AU11 + AU12 + AU25
Sadness	AU1 + AU4 + AU7 + AU15
Surprise	AU1 + AU2 + AU5 + AU26 + rotate-jaw

- *Anger (hot).* Tense voice, faster speech rate, higher pitch, broader pitch range;
- *Anger (cold).* Tense voice, faster speech rate, higher fundamental frequency and intensity, tendency toward downward-directed intonation contours;
- *Joy*. Faster speech rate, raised pitch, broader pitch range, rising pitch pattern;
- *Fear*. Raised pitch, faster speech rate, broadened pitch range, high-frequency energy;
- *Boredom*. Slower speech rate, additional lengthening of stressed syllables, lowered pitch, reduced pitch range and variability;
- *Sadness (crying despair)*. Slower speech rate, raised pitch, narrowed pitch range, narrowed variability;
- *Sadness (quiet sorrow)*. Slower speech rate, lowered pitch, narrower pitch range, narrower variability, downward-directed intonation contours, lower mean intensity, less precision of articulation;
- *Depression*. Lower intensity and dynamic range, downward intonation contours.

Looking at both lists, it is worth noting that there are similarities among the vocal expressions of moods and emotions (Isbister 2006: 186). For example, such different emotions as fear, anger, and joy are all signaled by faster speech rate, higher pitch, and increased pitch range. On the other hand, (quiet) sadness, boredom, and depression have a slower speech rate, lower pitch, and reduced variability in common. In these cases, additional cues may have to be provided in the voice, the content of the speech, and the speaker's facial expressions, gaze patterns, gestures, postures, and way of moving, in order to clarify what emotion is being expressed. Knowing another person well is also helpful because it includes knowledge about the particular ways in which that person expresses certain emotions through what he or she says, and how, as well as his or her visible displays of emotion (Isbister 2006: 186). This has two implications for designing the emotion expression of embodied interactive agents. First, while both the face and the voice are powerful means to convey emotions by themselves, an agent may often have to integrate them with each other and with further cues to convey an emotional message. Second, idiosyncratic ways of expressing emotions are part of an individual and hence should be designed into an agent. But the agent also has to do the best it can to help the user to get used to its expressive patterns, in particular if the user and the agent are to become companions for the long term, as in the relationship between a learner and an embodied pedagogical software agent in the role of fieldwork informant (cf. Chapter 16.2.2 and Chapter 16.3.5.1.3), coach (cf. Chapter 16.2.3), or learning companion (cf. Chapter 16.2.4).

Concerning what kinds of emotion should be expressed, the psychological literature identifies two general principles governing the expression of emotions in humans: hedonic preference and contextual appropriateness (Gong 2007). *Hedonic preference* means that people

Table 25. The FACS facial action units. Adapted from Pelachaud (2003, Table 5).

Action Unit	Name	Action Unit	Name
AU1	Inner Brow Raiser	AU31	Jaw Clencher
AU2	Outer Brow Raiser	AU32	Lip Bite
AU4	Brow Lowerer	AU33	Cheek Blow
AU5	Upper Lid Raiser	AU34	Cheek Puff
AU6	Cheek Raiser & Lid Compressor	AU35	Cheek Suck
AU7	Lid Tightener	AU36	Tongue Bulge
AU8	Lips Toward Each Other	AU37	Lip Wipe
AU9	Nose Wrinkler	AU38	Nostril Dilator
AU10	Upper Lip Raiser	AU39	Nostril Compressor
AU11	Nasolabial Furrow Deepener	AU41	Lip Droop
AU12	Lip Corner Puller	AU42	Slit
AU13	Sharp Lip Puller	AU43	Eyes Closed
AU14	Dimpler	AU44	Squint
AU15	Lip Corner Depressor	AU45	Blink
AU16	Lower Lip Depressor	AU46	Wink
AU17	Chin Raiser	AU51	Head Turn Left
AU18	Lip Pucker	AU52	Head Turn Right
AU19	Tongue Show	AU53	Head Up
AU20	Lip Stretcher	AU54	Head Down
AU21	Neck Tightener	AU55	Head Tilt Left
AU22	Lip Funneler	AU56	Head Tilt Right
AU23	Lip Tightener	AU57	Head Forward
AU24	Lip Presser	AU58	Head Back
AU25	Lips Part	AU61	Eyes Turn Left
AU26	Jaw Drop	AU62	Eyes Turn Right
AU27	Mouth Stretch	AU63	Eyes Up
AU28	Lip Suck	AU64	Eyes Down
AU29	Jaw Thrust	AU65	Walleye
AU30	Jaw Sideways	AU66	Cross-eye

generally prefer to perceive, experience, and express positive emotions over negative ones and therefore think that interaction partners displaying positive emotions are more attractive and appealing to work with (Gong 2007: 184). However, hedonic preference is constrained by the second principle, *contextual appropriateness*. That is, people usually adjust their emotional displays to make them appropriate for particular sociocultural contexts, following the display rules of their culture. Nass and Brave (2005) emphasized the importance of appropriateness of emotional tone for voices delivering different kinds of content. They argued that:

> *In sum, paralinguistic cues like speed, pitch, and volume are integrated with the perceived meaning of spoken words into a single message. Matching a voice to the content is just as*

Table 26. Vocal cues of basic emotions. Adapted from Nass and Brave (2005: 95, Table 8.1.A).

	Fear	Anger	Sadness	Happiness	Disgust
Speech Rate	Much faster	Slightly faster	Slightly slower	Faster or slower	Very much slower
Pitch Average	Very much higher	Very much higher	Slightly lower	Much higher	Very much lower
Pitch Range	Much wider	Much wider	Slightly narrower	Much wider	Slightly wider
Intensity	Normal	Higher	Lower	Higher	Lower
Voice Quality	Irregular voicing	Breathy chest tone	Resonant	Breathy blaring	Grumbled chest tone
Pitch Changes	Normal	Abrupt on stressed syllables	Downward inflections	Smooth upward inflections	Wide downward terminal inflections
Articulation	Precise	Tense	Slurring	Normal	Normal

> *important, and may even be more important than matching the voice to the emotion of the user. (Nass & Brave 2005: 88).*[226]

For cases in which the emotion of the content, the user, or the voice-based interface is not detectable or present, Nass and Brave suggested a slightly happy voice as a default in order to take advantage of people's hedonic preference (Nass & Brave 2005: 91). However, they also pointed out that expressions of anger or sadness tend to receive more attention and stick in the minds of the audience. Furthermore, cues of submissive emotions (e.g. sadness and fear) in the voice of a speaker are likely to increase the trust of listeners because they take these cues as indicators of the speaker being vulnerable and "opening up" to them (p. 92).

Embodied agents are able to express basic emotions through the manipulation of vocal cues as well as using the Facial Action Coding System outlined above. However, they still fall short when it comes to adapting emotion expression to changes within and across contexts:

> *[T]he state of technology today can easily enable agents including synthetic faces and speech to automatically generate effective positive or negative emotions. But it cannot yet automatically and reliably detect the emotional tone of the real-time interaction content and then correspondingly stipulate appropriate emotions to be displayed. In other words, without manual programming or simple and predictable scenarios, agents can easily fulfill the rule of hedonic preference but fall short in automatically meeting contextual appropriateness. (Gong 2007: 185).*

Gong (2007) conducted a study to investigate the question how far agents can go taking advantage of hedonic preference without achieving contextual appropriateness in emotion expression. A talking-head agent showing happy or sad emotional expressions presented novels with happy or sad content to the participants. Gong found that although neither 'happy' nor 'sad' agents adapted their expression to the emotional tone of the books, the agent showing the positive emotion elicited more positive responses to the books, the book reviews, and the agent and the interface. Furthermore, the overall user experience was more positive when the agent

[226] It is also important to maintain consistency between the content of the verbal message of an embodied agent and its emotional facial expressions, as demonstrated by Berry et al. (2005).

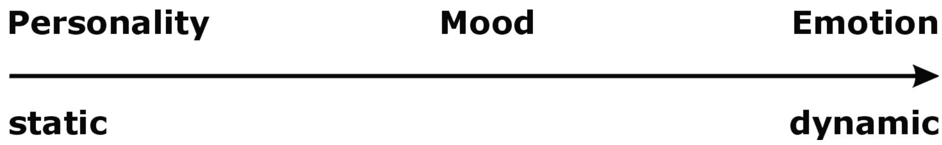

Figure 77. Personality, mood, and emotion on a scale from static to dynamic. Adapted from Egges et al. (2004, Figure 2).

appeared happy rather than sad. However, Gong was also careful to point out that "contextual appropriateness still appears necessary and important for achieving a satisfactory human-computer interaction experience" (Gong 2007: 189), given that the means of most of the measures in the study were fairly low, indicating that users did not like neither the happy nor the sad non-adaptive agent too much.[227]

13.5 Personality

Apart from how they feel at the moment about particular events, actions, or objects in their environment, the cognition and social interaction of individuals is also influenced by the attributes that define them as a distinct person. *Personality* is the complex of psychological traits that describe the unique character of an individual (Hayes-Roth et al. 1997), accounting for consistently chosen patterns of behavior, emotion and thought over situations and time (Moffat 1997; Ball & Breese 2000). Personality and emotion are essentially the same mechanism; however, they differ in terms of duration and focus (Moffat 1997). Personality is persistent and resistant to change over a lifetime, and it has a less clear focus than emotion (Gratch et al. 2002). In contrast, emotions are both short-lived and focus on particular events, actions, or objects (Moffat 1997, cf. Figure 77).[228] Personality influences how individuals interpret and respond to the environment, and it regulates their emotional state. Other things being equal, the behaviors of two individuals would show little difference were it not for differences in personality (Arnold 2003).

Personality is an essential property of life-like artificial agents because it functions as a *generative engine* that is instrumental in achieving individual appropriateness of an agent's emotional responses by contributing to their coherence, consistency, and predictability (Ortony 2003: 201f.). An agent that is (or portrays) a particular kind of person is influenced by its personality to respond to different kinds of situations in certain kinds of ways. In other words, being a particular kind of person (partially) determines the agent's behavior (p. 203).[229]

---

[227] Kim et al. (2007) found that pedagogical agents as learning companions (PALs) expressing positive emotions were perceived by learners as more facilitating to their learning, more engaging, and more intelligent. However, despite this demonstration of hedonic preference, Kim had previously made the case for contextual appropriateness by pointing out that "[t]o be effective, PAL affect should be tied to the learner's affect when possible. Rather than being simply a 'happy' talking head, a PAL should respond to or deal with the learner's affect and flexibly adapt its affect to the learner's in order to motivate the learner" (Kim 2005).

[228] A third category, *mood*, lies between these two extremes. Moods do not have a specific focus but are experienced as more diffuse, global, and general. They tend to have a longer lasting influence on an individual's cognitive strategies and processing (Brave & Nass 2003).

[229] Personality can thus be regarded as "the consistent expression of emotions" (Moffat 1997), and the personality of an agent can be created by building its emotional responses (Paiva et al. 1999).

Simple artificial personalities with a limited number of traits can be constructed trait by trait. However, with increasing diversity and complexity of traits and trait configurations, a more principled approach is required to create consistent and coherent personalities (Ortony 2003: 202). Matters are simplified by the fact that traits cluster together (Isbister 2006: 35)[230] and that two to five factors (depending on how they are grouped) are sufficient to describe the trait space (Ortony 2003: 203). Two popular five-factor models of personality (Wiggins 1996; John & Srivastava 1999) include the "Big Five" model (Digman 1990; Goldberg 1990; Goldberg & Saucier 1998) and the Revised NEO[231] Personality Inventory (NEO-PI-R), also referred to as the OCEAN model (Costa & McCrae 1992; McCrae & John 1992). Table 27 and Table 28 show the sets of factors defined by these models.

Embodied interactive agents project a personality both through how they portray themselves (using their face, body, and voice) and through what they do and how they do it. The design challenge involves creating a *personality expression* that uses all available communication channels to consistently and coherently convey the essence of a character using both verbal and non-verbal cues. Verbal cues include word choice, sentence types, fluidity of speech, ways of referring to other individuals, and use of indirect speech acts. Non-verbal cues include appearance (cf. Chapter 9.2), voice (cf. Chapter 10.1.1), gesture (cf. Chapter 10.1.2), facial features (cf. Chapter 10.1.3),[232] posture (cf. Chapter 10.1.5), and gait (cf. Chapter 10.1.6). In addition, different personality types are commonly associated with certain patterns of behavior (e.g. cautiousness vs. daringness). People observe all these cues and use them to interpret the personality of others (Nass & Brave 2005: 33). For example, there are observable differences in appearance and behavior between extrovert and introvert individuals, which are listed in Table 29.

The term 'personality' originates from the Latin verb *personare*, which means "to sound through" (a mask worn by an actor on stage). In fact, voice characteristics tend to play a more important role in assessing the personality of others than content of speech (Nass & Brave 2005: 33f.). Hence, it is not surprising that voice is one of the most effective ways to manifest a personality. Voices convey four types of cues that indicate personality (p. 34):

- *Volume* (how loud the voice is);
- *Pitch* (how high or low the voice is);
- *Pitch range* (how much the voice varies between high and low frequencies);
- *Speech rate* (how fast the person is speaking).

Nass and Brave (2005: 34) argued that while each of these aspects provides information about the personality of the speaker on its own, in concert, they have an even greater effect on the listener's judgment. For example, people are very likely to associate loudness, fast speech rate, high pitch, and wide pitch range with extroversion, whereas they would believe that a voice with the opposite characteristics belongs to an introvert person (cf. Table 29). Whatever the assessment will be, perceptions of personality in voice, looks, expression, and actions will

230 For example, friendliness is often found together with generosity, outgoingness, warmth, sincerity, helpfulness, etc. in a person (Ortony 2003: 202).

231 The acronym NEO consists of the initial letters of the original three factors in the model: neuroticism, extraversion, and openness.

232 In faces, dominance is signaled by strong features, a raised head, and lowered eyebrows. In contrast, a tilted head and raised eyebrows suggest submissiveness. A friendly person has a rounded face and large eyes. Pointed features indicate a character that is both hostile and dominant (Nass & Brave 2005: 139).

Table 27. The "Big Five" model of personality. Adapted from Pettijohn (1998: 281).

Factor	Example Traits (Positive Dimension)
Extroversion (or Surgency) vs. Introversion	Sociable, talkative, active, bold, fun-loving, spontaneous, adventurous, enthusiastic, person-oriented, assertive
Agreeableness (or Friendliness) vs. Antagonism	Warm, generous, trustful, courteous, agreeable, cooperative, flexible, forgiving, cheerful, humble
Conscientiousness (or Dependability) vs. Negligence	Conscientious, practical, cautious, serious, reliable, organized, careful, dependable, hardworking, ambitious
Emotional Stability vs. Neuroticism	Relaxed, peaceful, objective, calm, unemotional, even-tempered, secure, patient, uninhibited
Openness (or Culture, Intellect) vs. Closed to New Experiences	Original, imaginative, creative, perceptive, sophisticated, knowledgeable, cultured, artistic, curious, analytical, liberal

influence how a person feels and behaves with respect to other individuals. In particular, people will be drawn to others whose (voice) personality they perceive as *similar* to their own (cf. Chapter 10.1.1), perceiving them as more intelligent, trustworthy, and friendly, among other positive attributes (Nass & Brave 2005: 34).[233] As a result, the voice personality of an embodied interactive software agent should be cast with care. When using synthetic speech (cf. Chapter 11.1.2), the voice parameters relevant to personality (volume, pitch, pitch range, and speech rate) can be manipulated to induce different perceptions of personality (pp. 36ff.); however, designers should be careful to keep the personality projected by an agent's voice consistent (cf. Chapter 7.5) with the one projected by the other aspects of the agent's embodiment (cf. Chapter 9) and behavior (cf. Chapter 10). Nass and Brave argued that apart from similarity attraction, people exhibit *consistency attraction* with respect to personality. Whereas consistent personality cues guide people's expectations of and behavior toward others (Nass & Brave 2005: 33), perceived inconsistency of personality leads to feelings of confusion, uncertainty, and disturbance (p. 51).

Assessing the user's personality in order to adapt the agent's projected personality may involve direct methods in which users are asked to answer questions about their personality, or indirect methods that are based on analysis (e.g. of user profiles), inference (e.g. inferring personality type from information about the user's line of work), and user monitoring (Nass & Brave 2005: 40ff.). While visual cues of the user's appearance and behavior are still difficult to recognize and interpret for machines in general (cf. Chapter 11.2), agents can more easily identify characteristics of the user's voice to determine his or her personality. For cases in which the user's personality cannot reliably be detected, Nass and Brave (2005) gave the general recommendation of using an extrovert voice in an interface, which is more expressive (higher volume, faster speech rate, and increased pitch variation) than an introvert voice and

233 The principle of similarity attraction is only seemingly in conflict with the saying "Opposites attract." As argued by Nass and Brave (2005: 35), similarity attraction is actually responsible for opposites eventually liking each other. People have been shown to like others whose personality changes from being dissimilar to being similar to their own more than they like individuals who have been like them from the beginning (Moon & Nass 1996). This happens because "imitation is flattery," as another saying goes (Fogg & Nass 1997).

Table 28. The Revised NEO Personality Inventory (OCEAN model) (Costa & McCrae 1992).

Low Scorer Traits	Factor/Description	High Scorer Traits
	Openness (O)	
Conventional, down to earth, narrow interests, unartistic, unanalytical	Assesses proactive seeking and appreciation of experience for its own sake; toleration for and exploration of the unfamiliar	Curious, broad interests, creative, original, imaginative, untraditional
	Conscientiousness (C)	
Aimless, unreliable, lazy, careless, lax, negligent, weak-willed, hedonistic	Assesses the individual's degree of organization, persistence and motivation in goal-directed behavior. Contrasts dependable, fastidious people with those who are lackadaisical and sloppy	Organized, reliable, hard-working, self-disciplined, punctual, scrupulous, neat, ambitious, persevering
	Extroversion (E)	
Reserved, sober, aloof, task-oriented, retiring, quiet	Assesses quantity and intensity of interpersonal interaction, activity level, need for stimulation and capacity for joy	Sociable, active, talkative, person-oriented, optimistic, fun-loving, affectionate
	Agreeableness (A)	
Cynical, rude, suspicious, uncooperative, vengeful, ruthless, irritable, manipulative	Assesses the quality of one's interpersonal orientation along a continuum from compassion to antagonism in thoughts, feelings and actions	Soft-hearted, good-natured, trusting, helpful, forgiving, gullible, straightforward
	Neuroticism (N)	
Calm, relaxed, unemotional, hardy, secure, self-satisfied.	Assesses adjustment vs. emotional instability. Identifies individuals prone to psychological distress, unrealistic ideas, excessive cravings or urges and maladaptive coping responses.	Worrying, nervous, emotional, insecure, inadequate, hypochondriacal.

thus associated with predictability, which, in turn, is known to reduce listeners' cognitive load (cf. Chapter 2.2.2) and make them feel more comfortable. In addition, an extrovert voice may be perceived as friendlier and may receive more attention (Nass & Brave 2005: 42). However, Nass and Brave mentioned one exception to this extroversion rule, which is relevant to the design of pedagogical agents: when the system (or agent) should be perceived as *competent* (rather than likeable, cf. Chapter 7.2), a voice with deeper-than-average pitch should be chosen (p. 42).

Churchill et al. (2000: 80) proposed the following dimensions as a means to assess whether the design of the personality of an embodied interactive software agent is successful:

- *The lay personality psychologist test*. The agent's personality expression enables a human observer to create an informal description of the personality portrayed that provides a quite accurate picture of the agent's personality as intended by its designers (cf. the relationship between design model, system image, and user's model discussed in Chapter 6.3).

Table 29. Perceived characteristics of extroverts vs. introverts. Compiled from Nass and Brave (2005: 33f.).

Extroverts	Introverts
Charge into situations	Proceed cautiously
Talk more than they listen	Listen more than they talk
Speak loudly, rapidly, in a high pitch, and with a wide pitch range	Speak softly, slowly, with a deep and monotone voice
Use strong terms like *definitely* and *absolutely*	Use more neutral words like *perhaps* and *maybe*
Use many broad gestures	Use fewer, tighter gestures
Open their arms wide	Hold their arms close to the side
Hold their head erect	Hold their heads down
Tall and muscular	Short and frail

- *Shared traits*. Different embodied interactive agents may share certain personality traits (cf. Chapter 16.3.5.2), but the traits may be more or less pronounced in each agent.
- *Consistency vs. variability*. In general, the agent's responses to external events should be consistent with its personality (cf. Chapter 7.5). However, individual behaviors may involve a certain degree of randomness and unpredictability. Furthermore, repetitions of the same event may trigger slightly varied behavioral responses (cf. Chapter 7.9).
- *Different reaction tendencies*. The same event will elicit different responses from agents with different personalities. Human observers would explain the different behaviors in terms of the different personalities of the agents.
- *Perceptual filtering*. Personality serves as a filter through which the agent perceives external events. For example, a submissive agent may especially watch out for and respond to threats.
- *Modified presentation of self*. The agent can portray a modified personality expression in service of some goal. This change would be noticeable by an observer who knows the agent from previous encounters.
- *Adaptation*. During interactions with users, the agent can subtly adapt its personality expression in response to the personality or changing needs of its interaction partner.

13.6 Back-Story

What distinguishes people from machines is that behind every person, there is a rich individual *back-story*, which has been written since the day of their birth. Back-stories go beyond biographical data (family, education, job, etc.) to include idiosyncratic experiences, insights, memories, stories, and views. The closer we get to know someone, the more we learn about these things, which help us to understand and appreciate that person and make us share similar items in return. This mutual disclosure of personal information gradually establishes a bond between people (cf. Chapter 7.16), which may develop into friendship or even love.

Pedagogical agents should be designed to convey the impression of being an individual and to build relationships with learners. Both tasks are facilitated if the personality (cf. Section 13.5) of an agent is enriched with a back-story. Having its own experiences, insights, and stories to share with the learner allows the agent to enhance its believability as a person and facilitates establishing some common ground with the learner upon which a relationship can be built. The back-story of an agent may also contain items which are relevant to the current instructional situation and can be contributed by the agent to add flexibility and depth to its instructional conversations with the learner (cf. Chapter 16.3.5.3).

Writing the biography of another person is not an easy task, and it is even more difficult to invent the back-story of a synthetic agent which does not yet exist (and never will have an existence like a real person does). Striving for completeness is a hopeless endeavor because the description would soon lose itself in details. The writer needs to focus on creating a back-story that provides sufficient breadth and depth to support the agent's pedagogical, believability-enhancing, and interpersonal functions. The back-story should be internally organized in such a way that it is easy for the agent to relate the learner's contributions to items of the story and to retrieve relevant items that can be shared with the learner.

The story of a life is never finished until the person dies, so updating the back-story of an agent in a regular and timely fashion is an important issue. Again, it is not feasible to capture every small detail (unless the agent could be programmed to write its own life story), so the authors of back-stories have to make careful choices about what and when to update in order to maintain the illusion of a synthetic persona with a progressing life.

13.7 Summary

To become comfortable (in addition to useful and usable) for users, embodied interactive software agents should possess qualities that make them appear as emotionally intelligent individuals endowed with appealing and deep personalities. This chapter discussed the design of capabilities that allow agents to interpret, model, and adapt to the user's emotional state; generate appropriate emotional states within themselves; communicate ('genuine' or pretended) emotions to the user; and portray a rich and appealing personality and back-story.

To provide a background for this discussion, the nature of emotions was discussed in Section 13.1. After clarifying what emotions are, Section 13.2 was concerned with their functions in cognitive processes and social interactions. Two types of functions were identified: cognitive (intra-agent) functions (cf. Section 13.2.1) mediate various cognitive processes in an agent, whereas social (inter-agent) functions (cf. Section 13.2.2) influence social interactions between agents.

Equipped with the understanding of the nature and functions of emotions in humans, Section 13.3 covered the modeling of emotion in computers. Communication-driven approaches (cf. Section 13.3.1) select emotional displays for agents which are useful to achieve communicative, social, or task-related goals. In contrast, simulation-based approaches (cf. Section 13.3.2) aim to give an agent 'genuine' emotions by modeling the internal cognitive processes that give rise to observable emotional behavior. Three simulation-based approaches to emotion modeling were discussed in Section 13.3: the appraisal-based OCC model (cf. Section 13.3.2.1), one of its extensions, the Emotion and Adaptation (EMA) model (cf. Section 13.3.2.2), and an alternative approach based on Bayesian networks (cf. Section 13.3.2.3).

Section 13.4 was then concerned with emotional intelligence, i.e. the ability of an agent to recognize, interpret, predict, express, and manage emotions, moods, etc. within itself, the user, and other agents. Two component processes of emotional intelligence were discussed in this section: emotion interpretation (cf. Section 13.4.1), which infers the emotional state of the user, other agents, or the agent itself, and emotion expression (cf. Section 13.4.2), which generates emotional displays that are believable, easy to interpret, and observe the relevant display rules.

The final two sections covered aspects that define an agent as a person set apart from others. Section 13.5 discussed the design of personality, i.e. the complex of persistent psychological traits that characterize the agent as a unique individual. Two popular five-factor models of personality were presented (cf. Table 27 and Table 28), followed by a discussion of verbal and non-verbal cues for personality expression. Finally, section 13.6 argued that agents designed to convey the impression of an individual and to build relationships with users should be equipped with a rich, developing personal back-story that gives others an insight into their (scripted) lives. Writing the (ongoing) detailed back-story for an agent is a difficult task, so a number of suggestions were provided that may facilitate authoring these stories.

14 Architectures and Platforms

Whereas the previous chapters were concerned with the design of the different components of embodied interactive software agents, the focus of this chapter is on their underlying technical framework. Section 14.1 reviews various architectures that have been proposed to implement the internal mechanisms responsible for determining the actions and future internal states of an agent given its sensor input, as well as the collaborative problem-solving behavior of multiple agents.

Beyond the capability to act and solve problems, embodied interactive agents have to be able to generate believable behaviors in the context of real-time interactions with the user. Various approaches to behavior generation have been used for embodied agents, which provide different degrees of efficiency and flexibility as well as support for single or multiple characters. These approaches are discussed in Section 14.2.

Furthermore, embodied interactive agents always operate in some kind of computational environment and are therefore constrained by that environment, in particular by the qualities and deficiencies of the underlying technical platform. Different platforms for software agents are compared in Section 14.3. Finally, Section 14.4 summarizes the chapter.

14.1 Agent Architectures

A key problem for agents situated in a particular environment which they can sense and act upon is how to decide which actions will be most effective in helping them to achieve their design objectives in their current situation (Wooldridge 1999: 30). Various software architectures for situated decision making have been proposed in the literature. These architectures define a collection of component modules and how these modules interact to determine the actions and future internal states of a software agent given its sensor input data, as well as how groups of agents cooperate to solve problems beyond their individual capabilities. The following subsections review architectures for different types of software agents, including deliberative, reactive, hybrid, blackboard, and multi-agent approaches.

14.1.1 Deliberative Agents

Deliberation is the process by which a software agent weighs and chooses between alternative courses of action to achieve its goals. As shown in Figure 78, deliberation involves coming up with a set of alternatives and then picking one of them. The agent may consider both ends and means, i.e. both whether to attempt to achieve some goal and how to achieve it (Logan 2008c).

Deliberative agents have to maintain internal representations of both their environment and possible courses of action. Their reasoning processes manipulate these representations (rather than the environment) and produce descriptions of individual actions or action sequences to be carried out by the agent (Logan 2008c). Usually, these descriptions take the form of plans (cf. Chapter 12.1.3.3). Agents can generate plans to determine their future courses of action, by using a planning system with the following components (Jennings et al. 1998: 10f.):

- A symbolic model of the agent's environment (usually written in a manageable subset of first-order predicate logic, cf. Chapter 12.1.2.1);

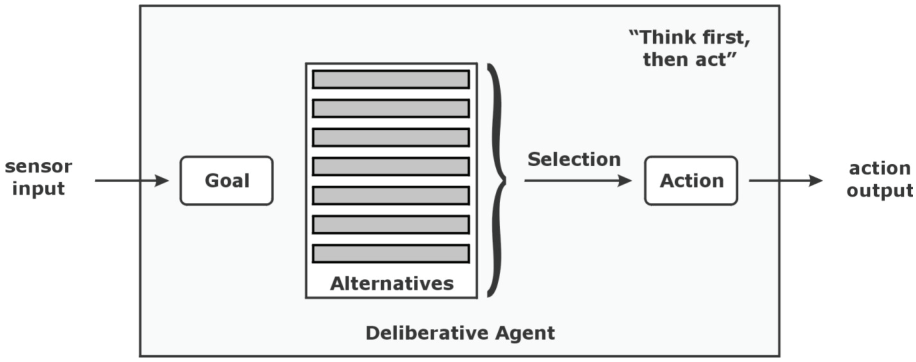

Figure 78. The architecture of a deliberative software agent. Adapted from Logan (2008c).

- A symbolic specification of the agent's inventory of actions (typically a set of plan operators with associated pre-conditions and effects);
- A planning algorithm, which computes a plan (or program) for the agent to achieve some goal, using the elements above plus a representation of the desired goal state as input.

Practical reasoning is a form of deliberation by which an agent decides on actions that serve to achieve its goals (cf. Chapter 12.1.3.4). The most widely used model of practical reasoning agents is the *belief-desire-intention (BDI) model* (Bratman et al. 1988; Georgeff et al. 1998; Wooldridge 2002: Chapter 4). The major components of a BDI agent are shown in Figure 79. They include (Wooldridge 1999: 57):

- A *belief revision function*, which computes new beliefs for the agent on the basis of the agent's perceptual input and its current beliefs;
- A set of *beliefs*, i.e. the knowledge that the agent has about its current environment;
- An *option generation function*, which identifies what options the agent has (i.e. its desires), based on its current beliefs and intentions;
- A set of *desires*, i.e. current options representing courses of actions that the agent could take;
- A *filter function*, i.e. the deliberation process which computes the agent's intentions from its current beliefs, desires, and intentions;
- A set of *intentions*, which represent the states of affairs that are the agent's current focus, in other words, which the agent has committed to trying to achieve;
- An *action selection function*, which decides what action to carry out given the agent's current intentions.

In general, deliberative agent architectures may be considered for applications in which mistakes made by agents are costly, and for implementing general procedures that solve whole classes of problems (rather than just one problem) (Logan 2008c). Their disadvantages include the difficulties of deriving accurate environment models and the high computational complexity of the underlying (planning) algorithms that makes real-time performance virtually impossible (Logan 2008d).

14.1.2 Reactive Agents

Deliberative agent architectures, like most work in artificial intelligence (AI) to date, are based on the symbol system hypothesis, which contends that intelligent behavior of machines can arise from the manipulation of structures composed of symbols representing objects or concepts

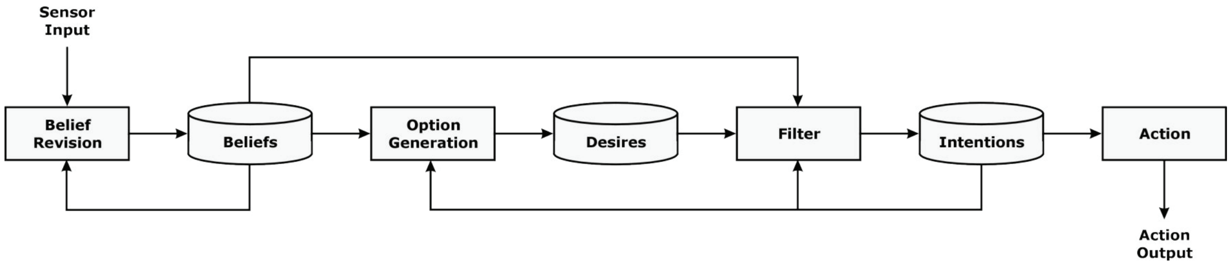

Figure 79. A generic belief-desire-intention architecture. Adapted from Wooldridge (1999: 32, Figure 1.5).

in the world (cf. Chapter 12.1.2). Since the second half of the 1980s, a growing number of AI researchers have distanced themselves from the traditional (symbolic) approach to building intelligent systems (Brooks 1986; Agre & Chapman 1987; Brooks 1990; Brooks 1991) and have formulated their own assertions about the nature of intelligence (Jennings et al. 1998: 12; Wooldridge 1999: 48f.):

- Intelligent behavior does not require explicit internal world models and complex deliberation processes.
- Intelligent behavior arises from the interaction between agents and their environments. It is not a characteristic of expert systems, theorem provers,[234] and other disembodied software systems that by the standards of traditional AI would be labeled 'intelligent.'
- Intelligent behavior is an emergent product of the interaction of simpler behaviors.

As an alternative to the symbol system hypothesis of 'classical' AI, the 'nouvelle' or 'fundamentalist' AI movement proposed the *physical grounding hypothesis*, which states that "to build a system that is intelligent it is necessary to have its representations grounded in the physical world" (Brooks 1990). Physically-grounded AI systems manipulate and respond to the environment itself rather than work with an internal model of it, as symbol-based systems do (Sengers 1998: 5). Taking the stance that "the world is its own best model" (Brooks 1990), nouvelle AI creates systems which often lack an internal symbolic environment model (Nwana 1996). A recurring theme in nouvelle AI is that of developing simple individual behaviors and combining them bottom-up into more complex behaviors, an idea that is at the heart of Brook's well-known *subsumption architecture* for mobile robots (Brooks 1986; Brooks 1990; Brooks 1991; Sengers 1998: 31f.; Russell & Norvig 2003: 932f.). Furthermore, intelligent systems typically are not regarded as disembodied entities but rather as situated (embedded) in some environment, connected to it via sensors and actuators (Brooks 1990; Sengers 1998: 6, cf. Chapter 3). Finally, because nouvelle AI systems tend to respond immediately to stimuli from the environment in which they are situated (Nwana 1996), without reasoning (too much) about its current state, they are sometimes called *reactive* (Wooldridge 1999: 49).

Reactive agent architectures are based on the physical grounding hypothesis and the nouvelle AI paradigm (Nwana 1996). In simple reactive architectures (cf. Figure 80a), sensor inputs are directly mapped to actions by means of condition-action rules (`if (sensor input) then action`) (Logan 2008b, cf. Chapter 12.1.2.4). Multiple inputs may trigger multiple actions, which may be performed in parallel (cf. Figure 80b), may be combined to create one

234 A theorem prover is a computer program for the automated proving (i.e. demonstrating the truth) of mathematical statements (or theorems) ([INT 159]).

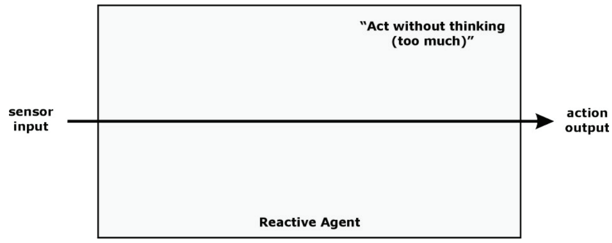

Figure 80a. The architecture of a simple reactive software agent. Adapted from Logan (2008a).

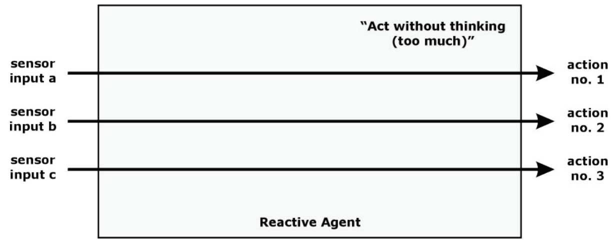

Figure 80b. Multiple parallel sensor inputs trigger parallel action outputs. Adapted from Logan (2008a).

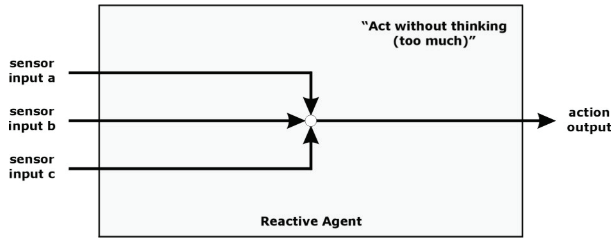

Figure 80c. Multiple triggered actions are combined into a single action output. Adapted from Logan (2008a).

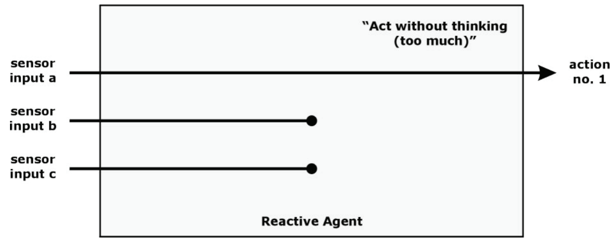

Figure 80d. One action output takes precedence over the other potential action outputs. Adapted from Logan (2008a).

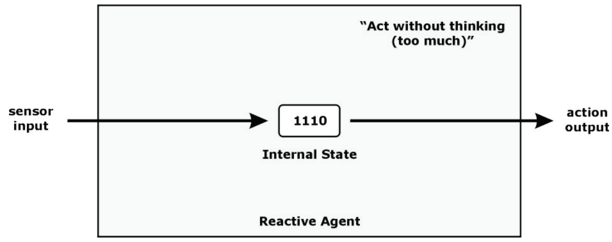

Figure 80e. A reactive agent architecture with a representation of internal state. Adapted from Logan (2008b).

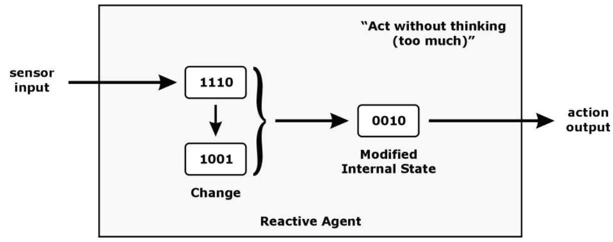

Figure 80f. A sensor input triggers a change of internal state, resulting in a modified internal state and an associated action output. Adapted from Logan (2008b).

composite action (cf. Figure 80c), or one of the triggered actions is executed with priority over the others (cf. Figure 80d) (Logan 2008a).

More complex reactive agents may additionally maintain an internal representation of their own state and/or the state of their environment (cf. Figure 80e), which enables them to select actions based on sequences of states, for example to respond to change or lack of it (cf. Figure 80f) (Logan 2008b). Such agents additionally require rules that are triggered by the internal state as well as rules with the ability to modify the internal state (Logan 2008b).

In principle, reactive agents maintaining internal state can generate any kind of behavioral response to sensor inputs in a timely fashion. However, both building complex internal models and reasoning about alternatives in problem solving are beyond the capabilities of these agents. Further problems concern the huge size of reactive agent programs for complex problems and the need to pre-specify all solutions to a given problem (Logan 2008b).

14.1.3 Hybrid Agents

There are many problems for which neither deliberative nor reactive agent architectures may be the most suitable approach, but where a combination of the two may succeed (Jennings et al. 1998: 13; Sycara 1998: 82). Hence, interest in *hybrid* agent architectures has grown, which seek to combine the strengths of both paradigms by integrating deliberative and reactive components (Jennings et al. 1998: 13; Logan 2008e).

Hybrid architectures are typically realized as a hierarchy of interacting software layers (Wooldridge 1999: 61), from lowest to highest, that process information about the environment at increasing levels of abstraction (Jennings et al. 1998: 13). Most hybrid agents use three layers (Sycara 1998: 82; Logan 2008e):

- A *reactive layer*, which quickly decides on basic actions in response to raw sensor input;
- A *deliberative layer*, which maintains a symbolic representation of the environment and determines and sequences actions for the reactive layer;
- A *social layer*, which regulates the social aspects of the agent's operation, in particular its coordination with other agents.

The integration of the layers is a central concern for the design of hybrid agents because the layers may use different representations and operate over different timescales (e.g. (fractions of) seconds for the reactive layer vs. minutes to hours for the deliberative layer) (Logan 2008e). The control mechanism used can be hierarchical or concurrent in nature (Jennings et al. 1998: 13; Logan 2008e):

- *Hierarchical control*. Higher-level layers control the operation of lower-level layers. Only the lowest level can access sensor inputs and action outputs (cf. Figure 81a).
- *Concurrent control*. The layers operate at the same time, as independent communicating processes. Sensor inputs and action outputs are accessible to all layers (cf. Figure 81b).

14.1.4 Blackboard Agent Architectures

In blackboard agent architectures (cf. Figure 82), multiple independent specialist agents (*knowledge sources*) cooperate to solve a problem, using a shared global workspace (the *blackboard*) for incrementally developing the solution, guided by a *control component* (Engelmore & Morgan 1988; Corkill 1991; Corkill 2003; [INT 160]).

Problem solving starts with a problem and initial data on the blackboard (Corkill 1991). Each knowledge source agent has special knowledge and capabilities for solving certain portions of the overall problem (Neumann 2000). When an agent's *triggering condition* is satisfied by information on the blackboard (usually events and/or data of a particular kind in which the agent is interested, generated by the agent itself or by other agents), the agent applies a *filter* to determine if this information allows it to make a contribution (Neumann 2000; Corkill 2003). If that is the case, the information is read from the blackboard, processed within the agent's *body*, and finally written back to the blackboard (Neumann 2000). In turn, this event triggers other agents, which likewise post their updates to the blackboard. This pattern is repeated until a solution is found ([INT 160]). A separate *control agent* manages the course of problem solving, selecting the next contributor(s) from a set of candidates based on the potential overall benefit of their contributions (Corkill 2003).

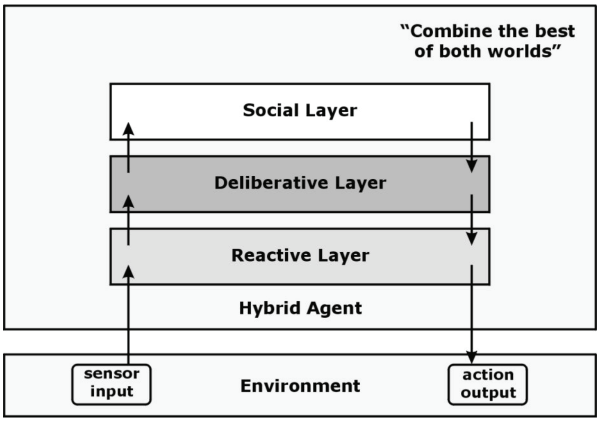

Figure 81a. Hybrid agent architecture with hierarchical control. Adapted from Logan (2008e).

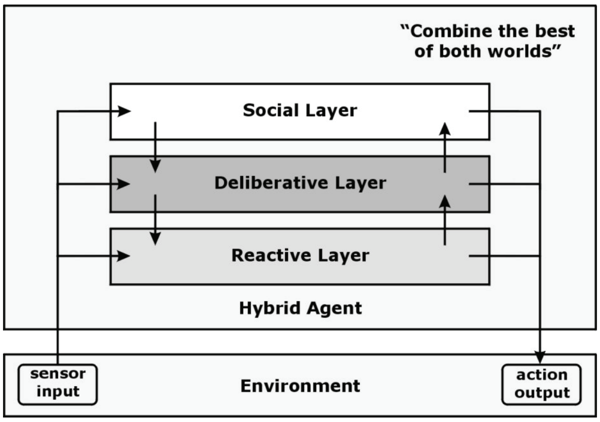

Figure 81b. Hybrid agent architecture with concurrent control. Adapted from Logan (2008e).

Blackboard architectures are suitable for ill-defined, complex problems involving diverse knowledge representations, problem-solving techniques, and developers ([INT 160]). They divide expertise into independent agents, which may use different technologies and can be easily added, modified, or removed from the system (Corkill 2003). The blackboard can represent any information; however, a common interaction language is required (Corkill 1991).

14.1.5 Multi-Agent Systems

Multi-agent systems originate from research on *distributed artificial intelligence*.[235] A *multi-agent system (MAS)* consists of a network of multiple interacting agents situated in the same environment (cf. Figure 83) that collaboratively solve problems which cannot be solved by the individual agents on their own (Stone & Veloso 1997; Sycara 1998; Weiss 1999; Wooldridge 2002; Vlassis 2007). Multi-agent systems have the following properties (Sycara 1998: 80):

- The agents in MASs are autonomous and may be of different types.
- The agents are designed to interact with one another, which involves cooperation (to achieve common goals), competition (to achieve individual goals at the expense of other agents), and negotiation (to resolve conflicts and to reach joint decisions).
- Individual agents in the network do not see the whole picture because they lack complete information or capabilities.
- Control and data are decentralized (i.e. distributed over the component agents).
- Computation is asynchronous (i.e. not happening at the same time).

Recent years have seen an increasing interest in multi-agent systems. This is mainly due to the following reasons (Sycara 1998: 80):

---

[235] *Distributed artificial intelligence (DAI)* is the subbranch of artificial intelligence that studies, builds, and applies systems consisting of several intelligent problem-solving entities that interact to achieve goals or perform tasks (Huhns 1987; Gasser 1989; Hewitt & Inman 1991). DAI has been largely supplanted by the field of multi-agent systems.

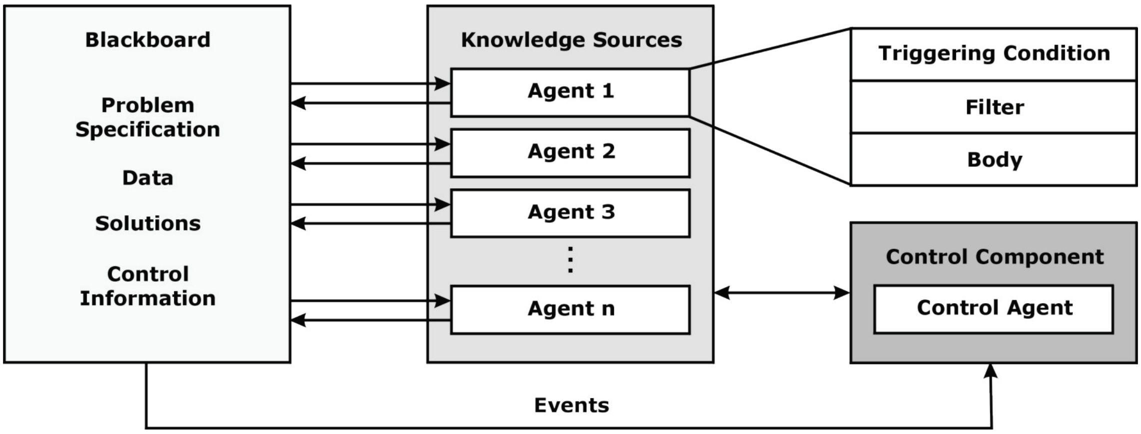

Figure 82. The blackboard agent architecture. Adapted from Corkill (1991, Figure 1) and Neumann (2000).

- Some application domains (e.g. meeting scheduling, air-traffic control, multi-agent bargaining for buying and selling goods on the Internet) require a multi-agent approach, in particular if they involve different people or organizations with different (possibly conflicting) goals and proprietary information.
- Multi-agent systems can solve problems that exceed the resources of a centralized agent or involve unacceptable risks of performance problems or failure if a single agent tried to deal with them on its own.
- Multi-agent systems can incorporate legacy systems (i.e. software systems that are important for an organization and have been in use for a long time), for example by placing the legacy software in an agent wrapper.
- Multi-agent systems can provide solutions in situations with distributed information sources (e.g. sensor networks, seismic monitoring, web-based information retrieval, etc.), expertise (e.g. health care, manufacturing, etc.), or control.
- A multi-agent approach can enhance the following aspects of system performance:
    - *Efficiency*. Multi-agent systems can speed up operation by assigning different tasks to different agents that operate concurrently.
    - *Reliability*. Multi-agent systems can effectively compensate for the failure of individual components by dynamically deploying free agent capacities.
    - *Extensibility*. It is possible to assign more and different agents to work on a problem.
    - *Robustness*. Multi-agents systems have the ability to compensate for uncertainty through information exchange between agents.
    - *Maintainability*. Components (agents) can be easily enhanced or replaced without affecting the others.
    - *Responsiveness*. Anomalies are handled locally and do not affect the entire system.
    - *Flexibility*. For different problems, agents can adaptively form teams that combine different capabilities.
    - *Reuse*. Agents with specific talents can find employment in different agent teams.
- Multi-agent systems provide a way to study intelligence by looking at the social behavior of societies of agents.

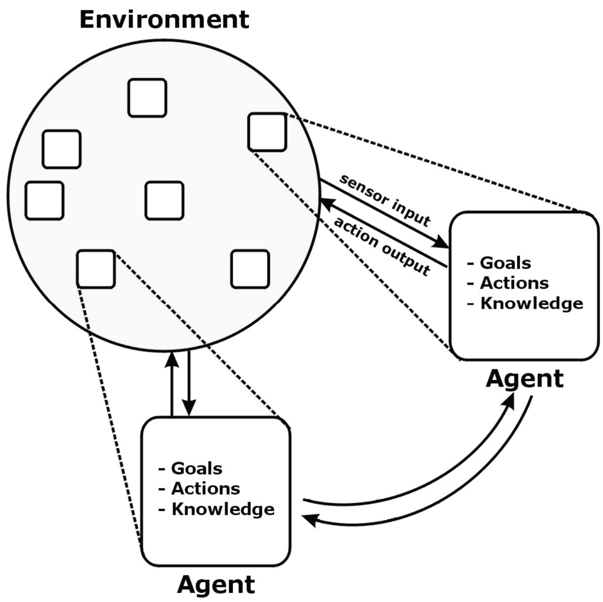

Figure 83. The architecture of a multi-agent system. Adapted from Stone and Veloso (1997, Figure 3).

While a promising concept, multi-agent systems may be difficult to design since the following issues, among others, need to be considered (Jennings et al. 1998: 17f.; Sycara 1998: 81; Vlassis 2007: 5):

- How to specify, decompose, and allocate problems for groups of agents and generate results from their interaction;
- How to enable inter-agent communication and with which languages and protocols; how to decide on the contents and timing of communication;
- How to enable each agent to model the actions, goals, and knowledge of its agent interaction partners;
- How to enable members of agent teams to represent and reason about the state of their coordinated problem solving;
- How to enable agents to identify and handle conflicts within and between groups of agents;
- How to enable negotiating and contracting among agents;
- How to reduce the danger or impact of harmful (chaotic or oscillatory) system behavior;
- How to form and dissolve teams, coalitions, and other structures that effectively accomplish assigned tasks or achieve specific goals.

In contrast to single-agent systems, the agents in a multi-agent system have to model each other's actions, goals, and knowledge (Stone & Veloso 1997). Coordination in multi-agent systems is achieved either by *direct communication* (i.e. managed by the agents themselves) or by *assisted coordination* (i.e. with the help of special *facilitator* components), with the agents exchanging messages in a common *agent communication language (ACL)* (Genesereth & Ketchpel 1994). Such a shared lingua franca is essential if the agents in a multi-agent system are heterogeneous (Sycara 1998: 87). It consists of (Wooldridge 2002: Chapter 8):

- A common *communication protocol* defining the syntax of messages. For example, the *Knowledge Query and Manipulation Language (KQML)* (Finin et al. 1994; Finin et al. 1997) provides such a standardized syntax, plus a set of performatives (e.g. 'tell,' 'perform,' and

'reply') that indicate the communicative force of messages (Wooldridge & Jennings 1995: 128).
- A common *interchange format* for message content. For example, the *Knowledge Interchange Format (KIF)* (Genesereth & Fikes 1992) provides a syntax based on predicate logic for specifying the content of messages (Wooldridge & Jennings 1995: 128).
- A set of shared *ontologies*, which define the entities, processes, etc. in the environment and how they relate to each other (cf. Chapter 12.1.2.5).

14.2 Behavior Generation

Beyond the capability to act and solve problems, embodied interactive software agents require the ability to generate coordinated verbal and non-verbal behaviors in the context of real-time human-agent interactions, which facilitate communication with the user, enhance the agent's believability, and help to establish human-agent relationships (cf. Chapter 10). The majority of current systems involving embodied interactive agents follow the *mind-body dichotomy* by distinguishing between an agent's embodiment and a higher level of behavior control (Rist 2004: 463). Technically, this distinction is often implemented as follows:

- *Embodiment*. An animation engine, often coupled with a speech synthesis system (cf. Chapter 11.1.2), generates the agent's digital representation (its embodiment, cf. Chapter 9), delivers the animations that portray the agent's visible behaviors, and produces the verbal utterances (spoken or written) that communicate the agent's messages. The component of an embodied interactive agent that is concerned with animation rendering and speech synthesis is often called the *character player* (cf. Chapter 3.3). Examples of freely available character player technologies include the Adobe (once Macromedia) Flash Player ([INT 161]), Microsoft Agent ([INT 162]),[236] and Cantoche Living Actor ([INT 163]).
- *Behavior control*. A behavior generation module formulates temporally ordered sequences of behaviors (gestures, gaze, facial expressions, postures, locomotion, spoken utterances, and bodily interactions with other actors and objects in the environment) in order to achieve communicative or task-related goals, often using planning (cf. Chapter 12.1.3.3) or other techniques from artificial intelligence or simulating aspects of human cognition. Behaviors have to be believable (cf. Chapter 3.1.3) and must be generated in real time. The output of the behavior generation component is a *script*[237] that is executed by the character player to produce observable behaviors performed by the agent persona (Rist 2004: 463).

Various approaches to behavior generation have been used in embodied interactive software agents, which provide different degrees of efficiency and flexibility as well as support for single or multiple characters. The following subsections discuss the most important architectures, including behavior sequencing, layered generation, state machine compilation, and multi-character systems.

[236] As of May 2009, Microsoft has stopped developing and supporting the Microsoft Agent technology ([INT 162]).

[237] Scripts define temporally ordered sequences of actions to be performed by an agent persona in order to realize communicative behaviors toward the user and other agents as well as bodily interactions with the agent's environment (Rist 2004: 463). Behavior can be scripted at different levels of granularity, ranging from low-level animation or voice commands to high-level specifications of complete activities that are hierarchically decomposed into a series of lower-level actions.

14.2.1 Behavior Sequencing

In the behavior sequencing approach (Stone & Lester 1996; Lester et al. 1999b; Lester et al. 2000; Lester et al. 2001), agent behaviors are composed from a library of pre-recorded primitive animations, sounds, and speech elements. These primitive elements are organized into a *behavior space* (cf. Figure 84), which is structured to assist in selecting and assembling agent behaviors that are aurally, visually, and semantically coherent (Stone & Lester 1996).[238]

At run time, a *behavior sequencing engine* selects elementary agent behaviors, audio clips of the agent's utterances, and other media elements from the behavior space and assembles them dynamically, constructing coherent paths through the behavior space in real time (Lester et al. 2001: 274f.). Variants of this approach are used by many embodied pedagogical agents, such as Vincent (Paiva & Machado 1998; Paiva & Machado 2002), Adele (Shaw et al. 1999; Johnson et al. 2003a, cf. Figure 17a), Herman (Lester et al. 1999a; Lester et al. 2001), and Cosmo (Lester et al. 1999b; Lester et al. 2001, cf. Figure 17c).

Behavior sequencing is appropriate for applications with a fixed camera position. Assembling behaviors out of prerecorded segments saves time and can produce high-quality animated behaviors, provided that the segments are created by expert animators (Johnson 1998). The major shortcoming of behavior sequencing is its inability to adapt behavior sequences to users' actions in real time. A single action of the user performed during the execution of a behavior sequence can render the whole sequence inadequate. Furthermore, creating a coherently structured behavior space is a complex, labor-intensive task (Johnson et al. 2000), which requires human resources with multiple talents (cf. Chapter 7.12).

14.2.2 Layered Generation

In contrast to behavior sequencing, the layered generation approach (Johnson et al. 1998; Rickel & Johnson 1999; Rickel & Johnson 2000) creates animated behaviors in real time rather than assembling them from a collection of pre-crafted multimedia elements. The architecture used consists of two layers: a *cognitive decision-making layer* and a *perceptual-motor layer* (cf. Figure 85).

From the various events occurring in the environment, the *perception module* identifies those that are relevant to the agent in order to provide a coherent view of the state of the environment to the cognition module. The task of the *cognition module* is to interpret the information provided by the perception module and, based on the agent's goals and internal state, to decide on the agent's next actions. It outputs high-level commands to the motor control component, which decomposes them into a series of lower-level motor commands that eventually produce animated and verbal agent behaviors in the environment (Rickel & Johnson 2000: 105f.).

The main advantage of the layered generation approach is the clear separation between agent decision making and the modules for sensing and motor control. Furthermore, a library of pre-fabricated primitive animations is not required; the agent's behaviors can be rendered on the fly using well-defined rules and can be flexibly controlled with motor commands down to the small details even while they are being executed, in particular to adapt the sequence of behaviors to some unforeseen changes in the environment (Rickel & Johnson 1999). The chief disadvantage

[238] The original publications on the topic wrote "pedagogically coherent" instead, but that expression seemed to be too specific in the context of the present discussion.

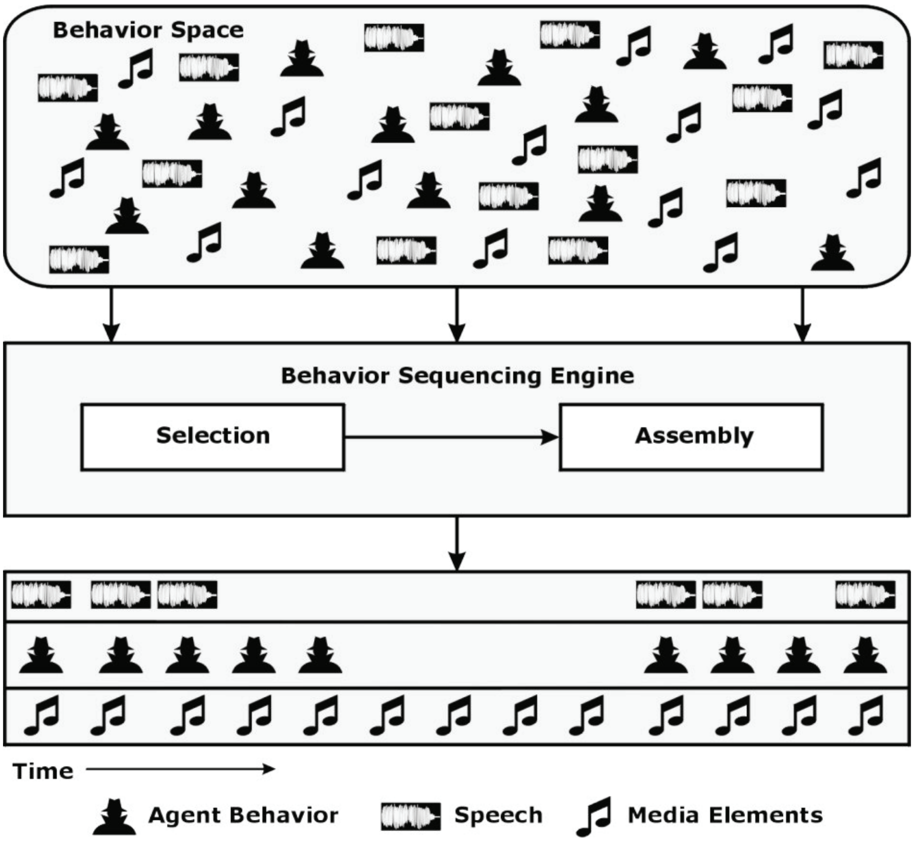

Figure 84. An overview of behavior sequencing showing the behavior space, the behavior sequencing engine, and the assembled behavior sequence. Adapted from Lester et al. (2001: 274, Figure 2).

of the layered generative approach is the increased amount of rendering computation required to create agent behaviors. Furthermore, the quality of the automatically generated animations is usually lower than the quality of animations created by human animators (Johnson et al. 2000).

14.2.3 State Machine Compilation

The state machine compilation approach (André et al. 1999b) seeks to achieve real-time adaptation of agent behavior at acceptable computational costs. The embodied agent acts as a presenter of multimedia information, using a combination of speech and gestures to refer to elements of the presentation.

The architecture of the presentation engine is shown in Figure 86. The approach taken is similar to behavior sequencing in that behaviors are composed from elementary agent postures that correspond to a single animation frame or an uninterruptible image sequence. However, the major difference is that for efficiency, behaviors are compiled into a state machine (the *behavior monitor*) that executes them at run time (Johnson et al. 2000: 73), combining planned presentation actions with unplanned self-behaviors, such as idle-time actions (e.g. breathing, toe tapping) and reactive behaviors (e.g. hanging suspended when the user drags the agent with the mouse), in order to adapt to users' actions and to increase believability (cf. Chapter 10.2) (Johnson et al. 2000: 69).

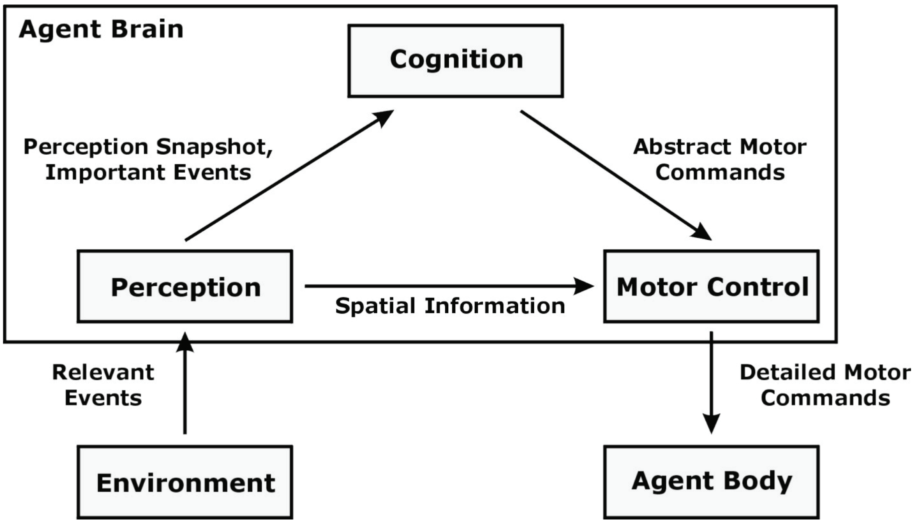

Figure 85. The layered behavior generation architecture. Adapted from Rickel and Johnson (2000: 106, Figure 4.5).

The state machine defines a space of possible agent behaviors. A presentation plan specifies the set of presentation actions to be performed plus a schedule for performing them. At run time, the agent gives its presentation following the preliminary schedule. When the behavior monitor carries out additional actions, the schedule may have to be updated, observing the constraints imposed by the presentation plan. The agent is thus able to perform sequences of actions that are both adaptive and interruptible but still as coherent as possible (Johnson et al. 2000: 69).

14.2.4 Multi-Character Systems

Multi-character systems (Rist et al. 2003; Rist et al. 2004) script and coordinate the behavior and interaction among and between multiple animated characters and human users. The MIAU[239] platform (Rist et al. 2003) is a multi-agent architecture (cf. Section 14.1.5) that is tailored to applications involving multiple characters (and users) in different conversational settings ranging from non-interactive presentations involving a single user and a single agent to multi-party multi-threaded conversations (see Chapter 12.3.12 for a discussion of the different scenarios). The elements of the MIAU platform are shown in Figure 87. They include a set of character components, a director component, and a user component.

Each *character component* C_1, ..., C_n controls an animated character and consists of a behavior planner and an interaction manager. The *behavior planner* decomposes complex discourse goals into basic acts for execution by the character player. The *interaction manager* keeps track of the interaction state and maintains a history of the interaction. It also handles the internal communication with other components by means of formalized communication acts that are exchanged through the shared *message board* (Rist et al. 2003).

The *director component D* controls the interactive performance as a whole. It contains a *narration planner* module which shapes the course of the performance. Different interaction styles are realized by (dynamically) shifting control between the director component and the

[239] MIAU is obviously an acronym, but unfortunately, none of the publications reviewed spelled it out.

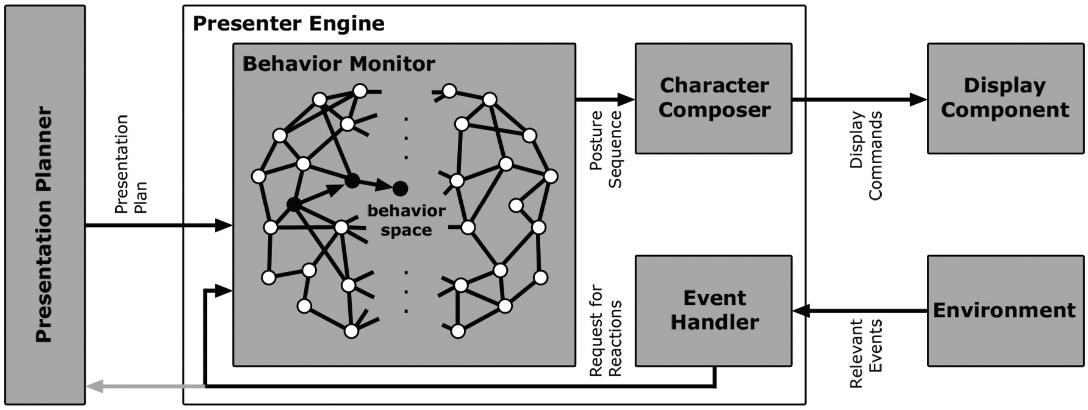

Figure 86. The architecture of the presenter engine. Adapted from André et al. (1999a, Figure 12).

character components. For example, the director may script all actions for the individual characters in detail, or the characters may have greater autonomy and the director intervenes only from time to time to increase coherence, introduce new topics, etc. The MIAU platform can flexibly switch between different conversational settings (character monologue, character dialogue, multi-character performances, and (multi-) character performances with user interaction) by adding and removing characters and avatars (representing the user, cf. Chapter 3.1.4) in the course of a performance (Rist et al. 2004: 393). The *interaction manager* handles the exchange of messages with other modules via the message board.

The *user component U* allows for user intervention, enabling the user to become an active participant in a character performance, for example by asking the participants questions, and even to influence the performance as a director or co-director (Rist et al. 2004: 392). The *input analyzer* maps from user inputs to internal communication acts. Input activity may be indicated by the user's avatar. For example, the avatar may speak the user's typed input (Rist et al. 2003). Again, the *interaction manager* handles internal communication. While Figure 87 shows a single user component, the architecture can be extended to support multi-user multi-character scenarios by increasing the number of user components (Rist et al. 2004: 393).

The major advantage of multi-character platforms like MIAU is the flexibility which they add to applications involving one or several embodied agent presenters. They can be used to implement a wide range of scenarios ranging from traditional single-presenter non-interactive presentations to multi-party multi-threaded conversations. In addition, MIAU allows dynamic switching between conversational settings or between director- and character-centered scripting approaches during a session. A third advantage of MIAU's architecture is that it separates the specification of scripting knowledge from the design of the computational mechanisms for behavior generation; in other words, there is a clear distinction between knowledge engineering (cf. Chapter 12.1.1.1) and implementation (Rist et al. 2004: 401).

However, multi-character systems also involve a number of issues. The scenarios envisioned for these systems become more and more complex as the number and interaction possibilities of characters and users involved increases. Scripting the interactions among several animated characters already involves considerable challenges with respect to creating, maintaining, and extending the structure and content of exchanges because manually pre-authored scripts may have to be integrated with scripts that are automatically generated at run-time, and different types of conversations between the virtual performers (e.g. 'on-duty' vs. 'off-duty') have to be supported (Gebhard et al. 2003). But the situation becomes even more difficult when the human user is allowed to participate in the performance, especially if he or she is to be given the freedom of expression that she has in natural face-to-face conversations, i.e. to combine verbal

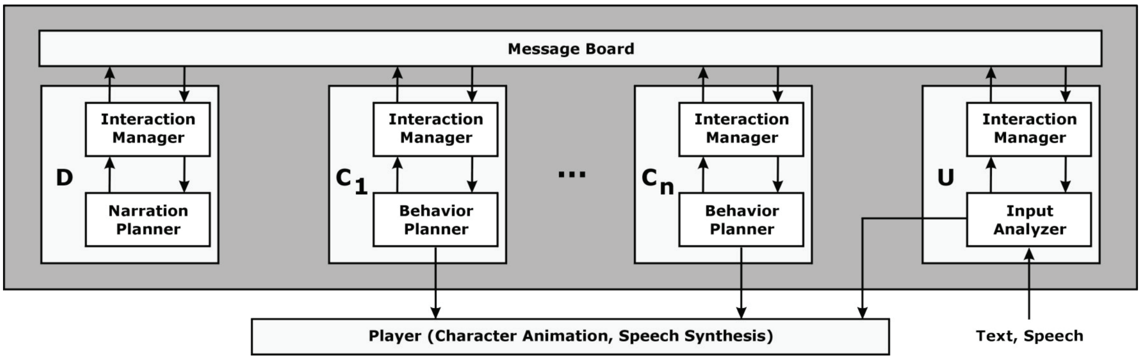

Figure 87. The MIAU multi-character platform. Adapted from Rist et al. (2004: 393, Figure 9).

and non-verbal modes of communication. The capacity for unrestricted natural language (or even multimodal) conversations between users and agents, while highly desirable, will remain unattainable for years to come, although various more restricted (embodied) conversational agents have become feasible (cf. Chapter 11.3). In the most ambitious setting, multi-party multi-threaded conversations, where several users and several characters can interact among themselves and with each other, the major challenge is to keep the performance on track while allowing the participants to interact freely (Rist et al. 2004: 378). In addition, the system must be able to follow the different threads of conversation and detect where they overlap. Further difficulties arise from the shortcomings of the technologies required to recognize and interpret what the human participants say to each other (p. 379).

Van Deemter et al. described a multi-character system that automatically generates whole scripted multimodal dialogues which are acted out by a cast of embodied conversational agents using combinations of text, expressive speech, and non-verbal communicative behaviors (van Deemter et al. 2008). In these character performances, the virtual actors have to play their parts in the dialogue convincingly, which includes the expression of different emotions (cf. Chapter 13.4.2). For this, the actors do not have to be equipped with the ability to interpret (spoken) language because their exchanges are pre-determined by a script that is (semi-) automatically produced from a high-level parameter specification;[240] however, their acting should create and maintain the illusion that they understand and respond to each other both at the conversational and at the affective level (van Deemter et al. 2008, cf. Chapter 16.3.3). Each generated dialogue should be (slightly) different from all others; in other words, the generation process should provide for variability (cf. Chapter 7.9) at all levels (van Deemter et al. 2008).

In the system described by van Deemter et al., a script for the scene to be performed by the virtual actors is generated first, which specifies the semantic and affective content of the dialogue and its organization as an XML (cf. Footnote 182) document that is written in the Rich Representation Language (RRL) (Piwek et al. 2002). This high-level scene description becomes more and more detailed in the following processing stages that generate appropriate and coordinated language, emotionally inflected speech, and non-verbal behaviors for the virtual actors participating in the dialogue. Van Deemter et al. stressed the incremental nature of their architecture, which never discards information but only adds more at each stage, thus allowing later stages to make use of all the information added so far (van Deemter et al. 2008).

[240] It should be noted that the situation changes as soon as the user is allowed to join the conversation. The course of the dialogue is then no longer predictable, and the system also has to deal with the user's linguistic inputs (cf. Chapter 12.3.12 and Chapter 16.3.4).

14.3 Technical Platforms

The technical platform of a software agent is the specific configuration of hardware and software components providing the framework for the agent's operation and the user's experience. It may consist of computers, networks, input and output devices, system and application software. All these components have their own capabilities and limitations and interact in complex ways.

Software agents have been developed for a variety of technical platforms, including desktop computers, the World Wide Web, virtual reality simulations, and mobile devices. These platforms are discussed in the following sections.

14.3.1 Desktop Computers

A desktop computer is a stand-alone personal computer designed for use by an individual user in a home or work environment. It may be connected to other computers through the Internet or some other computer network, but it does not depend on a network connection for its operation. It typically runs an operating system from Microsoft's Windows ([INT 164]) or Apple's MacOS family ([INT 165]), or an open source operating system like Linux ([INT 166]). Typical application software used on desktop computers includes office suites, web browsers, e-mail clients, and computer games. Today, desktop computers are increasingly complemented or replaced by laptop computers, i.e. small portable all-in-one computers explicitly designed for mobile use, which essentially have the same components and capabilities as desktop computers but offer users the advantage of mobility, whereas a desktop computer cannot easily be moved from one workplace to another.

Software agents embedded in a desktop computing environment perform tasks defined and enabled by the software on that computer for a well-defined range of users, i.e. those users having access to the computer system (and the agent). They may be part of a particular piece of software (e.g. for learning or entertainment) or their responsibilities may span multiple applications and/or system programs (cf. Chapter 3.2.5).

All components of a desktop agent run on the local computer and can take full advantage of its processing, storage, multimedia, and networking capabilities. The agent may be installed as a complete package, without the need to worry about missing plug-ins or function libraries. As a result, it is possible to equip the agent with advanced features, such as speech recognition (cf. Chapter 11.1.1) and speech synthesis (cf. Chapter 11.1.2).

However, developers usually have to provide implementations of their agents for different desktop platforms and may have to deal with compatibility issues involving software already running on the user's computer. Furthermore, there will not be one central but rather multiple distributed installations of an agent system on different hardware and software configurations, which might be difficult to debug and maintain (although desktop computers with an Internet connection can automatically receive updates online).

14.3.2 World Wide Web

The World Wide Web is a service for accessing and disseminating information that operates over the Internet, a massive global networking infrastructure connecting millions of computers. At the technical level, the World Wide Web is a *client/server system*, i.e. a networked system in which requesters (clients) and providers (servers) of services communicate over the underlying network(s) using request-response pairs as defined by the *Hypertext Transfer (HTTP) protocol*. In this network, processing is distributed among clients and servers. A number of issues need to be addressed when integrating software agents into web-based environments, which are discussed in the following.

14.3.2.1 Network Latencies

The most severe technical problem of integrating agents into web-based environments is to find workarounds for the *slow response times* of the World Wide Web. From face-to-face interactions in the real world, users will expect a web-based agent to be able to respond in (near) real time to a continuous stream of actions rather than discrete request-response pairs (Johnson et al. 2000: 73). Tolerance for delays caused by waiting for a server to create agent behavior in response to a user's action will usually be minimal. In addition, the agent may have to coordinate the activities of multiple agents and users on different clients, for example in a collaborative activity such as a team training scenario (p. 73).

Network latencies can be dealt with by finding an appropriate distribution of agent components between server and clients:

- Modules implementing the agent's persona and behaviors (both reactive and proactive) should run on the client or a dedicated server.
- The agent's domain knowledge (cf. Chapter 12.1) and user models (cf. Chapter 12.2.1) should remain on a central server. Relevant parts can be downloaded to the local client and processed locally by lightweight components.
- It is essential to maintain a balance between the workload of the client, the server, and the network that connects them.

14.3.2.2 Performance

Apart from the network connecting clients and servers, the performance of the computers in the network themselves, i.e. of the user's computer and the server machines, constitute another bottleneck. While necessary upgrades can be fairly easily performed on the servers (subject to technical and monetary restrictions), developers generally cannot control what hardware people buy and use. Consequently, the agent will have to perform reasonably well on a wide range of client machines with different processing, storage, multimedia, and networking capabilities. This constraint affects in particular the rendering of agent embodiments and their behaviors (cf. Chapter 9 and Chapter 10). High-quality animations of dynamically generated agent behaviors depend on powerful processors and graphics accelerators (Johnson et al. 2000: 72).

Given the heterogeneity of client computers, compromises are inevitable. For example, instead of creating agent behaviors dynamically, it may be more feasible to assemble them from pre-crafted elements (cf. Section 14.2). Another possibility is to adapt the range and detail of the agent's animations to the capabilities of particular platforms (Johnson et al. 2000: 72). Care has to be taken, though, that downgrading the agent impairs neither its believability nor its task-related and social functions (cf. Chapter 8.4).

14.3.2.3 Technologies

Integration and interoperation of a software agent with a web-based environment is facilitated if both make use of the same technologies to implement appearance and functionality. The major responsibility of client-side technologies is to create the appearance and behaviors of the web-based agent and to handle the interaction with the user. The following technologies are available in many current web browsers to implement the agent-user interface:

- Adobe Flash/Shockwave ([INT 161]) for agent appearance and animation;
- HTML ([INT 167]) for embedding the agent in web pages;
- JavaScript ([INT 168]) for interaction with users and the environment;
- Java ([INT 169]) for aspects of the agent that require more complex processing.

As a general recommendation, developers should refrain from relying on special software components (plug-ins) to make their agent work in the web-based environment. Any additional software which the user has to install is a potential source of error and an obstacle to working with the environment and the agent. Instead, the agent should run in a web browser equipped with just the common components, including those listed above. In general, relying on standard web technologies avoids the plug-in problem and facilitates integration and interoperability.[241] Lightweight processing components (applets, modularization) and formats (e.g. vector-based graphics) as well as *streaming*[242] help to minimize the load of the network connection and the client computer.

On the server, a collection of scripts and/or programs is required to manage those aspects of the agent and the environment that have been allocated to a central server, plus a database system to store information about the agent, the environment, and the users in a central place. Server-side components can be written in any programming language. A number of scripting languages and relational database systems are freely available and thus provide cost-efficient options to implement server-side processing for web-based environments with embedded software agents.

14.3.3 Virtual Reality

Virtual reality (VR) is a computer-generated, interactive, graphical (usually three-dimensional), or multisensory simulation of some environment (Rheingold 1991; Burdea & Coiffet 2003). The simulated environment may be similar to the real world (e.g. simulations for pilot or combat training), or it can differ significantly from reality, as in VR computer games.

Most current virtual reality experiences are created by *desktop VR systems*, which display the virtual environment on a conventional computer screen, now typically including sound through speakers or headphones and using standard input devices, such as a keyboard, mouse, and joystick. Limited tactile, haptic force feedback may also be available (cf. Chapter 1.2).

[241] However, this also means that agent development has to work around the various limitations of the standard web technologies, which would be less of a problem if a custom agent plug-in were used.

[242] Streaming transfers data in such a way that it can be displayed or otherwise processed continuously by the recipient while it is being transmitted. On the World Wide Web, streaming is typically used for audio, video, and other kinds of high-volume data.

In contrast, immersive VR systems move the user's viewpoint completely into the virtual world. Head-mounted displays give the user the impression of being embedded (or *immersed*) in this world. Interaction with the environment usually involves the use of manual input devices like data gloves.

Integrating embodied interactive agents into virtual reality simulations requires addressing a number of technical issues:

- Virtual reality is a demanding technology in terms of processing speed and graphical rendering capabilities. Complex, high-fidelity agents and environments may not be feasible on the user's computer. The implementation must be able to reduce detail levels and complexity gracefully to maintain the user's experience while adapting to the capabilities of his or her hardware and software.
- The size of the computer screen or the field of view of the head-mounted display limits what the user can see of a virtual environment. As the user moves closer to particular objects in order to manipulate them, they may lose sight of a fully embodied agent, or the agent's body may block their view. Therefore, agents in virtual environments must be able to move out of or into the user's field of view (cf. Chapter 10.1.6) or switch to an alternative embodiment (cf. Chapter 9.7) (Johnson et al. 1998).
- The agent needs to monitor the user's activities in the virtual world (cf. Chapter 4.2). In immersive VR systems, it is easier for the agent to locate users and determine what they are doing, where they are looking, etc. because that information is provided by the VR system itself. In contrast, desktop VR systems require additional devices (cameras, microphones, pressure sensors, etc.) to monitor the user.
- Agents immersed in virtual reality environments must enable multimodal interaction with the user (cf. Chapter 11). They have to be able to interpret combined user input modes (speech, gesture, etc.) and to effectively control their own virtual body (e.g. gaze/point at or manipulate objects). Agent development should consider that some input technologies (e.g. keyboard and mouse) are more advanced than others (e.g. face and gesture recognition), and enable the agent to flexibly use the most effective combination (cf. Chapter 11.2).
- Virtual environments may be populated by many human users and agents. The architecture of an agent must be designed in such a way that the agent is aware of and can coordinate the activities of multiple human and artificial participants in the virtual world. This presupposes the ability to distinguish between humans and agents, which may portray themselves in similar ways.

14.3.4 Mobile Devices

Mobile devices are pocket-sized computers that have become very popular as a convenient tool for people to store, organize, and access information, to entertain themselves, and to communicate with others anytime and anywhere. The reason for the ubiquity of these devices lies in the importance that people place on constant access to information, entertainment, and communication both in their professional and their private lives. Examples of mobile devices include cell phones, personal digital assistants, digital media players, and handheld game consoles.

Software agents can be embedded in mobile devices to manage calendars, access e-mail and voice mail, search the World Wide Web, select music and video clips, serve as an answering machine, play adversaries in computer games, etc. The implementation of agents for mobile devices needs to consider the following limitations of these artifacts (Chittaro 2006: 40):

- Mobile devices have smaller displays with lower resolution and fewer colors than the screens of desktop computers. As a result, the agent's embodiment has to be smaller and simpler. On devices with a very small display or none at all, it may not even be possible to use a body.[243]
- Keyboards are scaled down due to size limitations. Mobile devices therefore increasingly use pen-based or voice-based input, either as a complement or as a replacement for keyboards. Hence, the agent must be prepared to handle multimodal input, including voice and gesture (cf. Chapter 11.2).
- Memory, storage, and processing capacity are limited, and so is the speed of the network connection. Consequently, the agent should only use lightweight processing components and limit network access for its own purposes to the necessary minimum.

14.4 Summary

This chapter focused on the technical side of embodied interactive software agents, reviewing software architectures for situated decision making, approaches to behavior generation, and technical platforms for software agents. Section 14.1 compared five different architectures for implementing the internal mechanisms that allow an agent to determine its actions and future states given its sensor input: deliberative (cf. Section 14.1.1), reactive (cf. Section 14.1.2), hybrid (cf. Section 14.1.3), blackboard (cf. Section 14.1.4), and multi-agent (cf. Section 14.1.5) architectures.

The second part of the chapter was concerned with the problem of generating the behaviors of an embodied interactive agent. Typically, solutions to this problem involve two components: a character player for producing the agent's animations and utterances and a higher level of behavior control that generates scripts specifying temporally ordered sequences of actions which are executed by the character player. Approaches to behavior generation differ with respect to their efficiency and flexibility as well as their support for single or multiple characters. Four of these approaches were reviewed in Section 14.2: behavior sequencing (cf. Section 14.2.1), layered generation (cf. Section 14.2.2), state machine compilation (cf. Section 14.2.3), and multi-character systems (cf. Section 14.2.4).

Software agents are designed for a particular computational environment, which consists of a specific configuration of hardware and software components that exhibit individual capabilities and limitations as well as complex interactions with each other. Section 14.3 discussed the characteristics of four technical platforms for embodied interactive software agents: desktop computers (cf. Section 14.3.1), the World Wide Web (cf. Section 14.3.2), virtual reality simulations (cf. Section 14.3.3), and mobile devices (cf. Section 14.3.4). Various technical issues were identified that need to be addressed when developing agents for these platforms, such as the slow response times of the World Wide Web or the small display sizes and limited input and processing capabilities of mobile devices.

[243] Small display sizes are not necessarily a disadvantage. Johnson et al. (2004c) argued that on a small screen, glitches in character animation are less noticeable. Furthermore, users may generally expect lower production quality on handheld displays. Finally, if the agent has a smaller and simpler body, the task of the character animators becomes easier.

15 Evaluation

[T]he simple question as to whether an animated interface agent improves human-computer interaction does not appear to be the appropriate question to ask. Rather, the question to ask is: ***what kind of animated agent used in what kind of domain influences what aspects of the user's attitudes or performance?*** *(Dehn & van Mulken 2000: 19, their emphasis).*

The increasing interest in embodied interactive software agents for applications in e-learning (cf. Chapter 2) and other domains largely originates from the assumption that these agents have positive effects on human cognition and motivation by making computer-based systems more human-like, more engaging, and more motivating (De Angeli et al. 2001b). While anecdotal evidence in the form of informal observations, reports, or user feedback may or may not support this assumption, it typically fails to inform designers about exactly which features of (a particular specimen of) embodied interactive agents have what effect on what group of users in what situations and for what purpose. Designers have to make many choices when building an agent playing a specific role in a specific context for the target group of users. These design choices and their interactions have been discussed in detail in Chapter 9 to Chapter 13. Often, choices made in agent design lack a solid foundation. Isbister and Doyle criticized that:

Choices about the appearance, personality, and behaviors of the agent are frequently made on the basis of an introspective examination of personal preferences, and in many cases do not accurately reflect the goals of the design or the qualities of the audience with whom the agent is ultimately intended to interact. (Isbister & Doyle 2002).

The question addressed in this chapter is how to determine effective combinations of agent design parameters (cf. Section 15.2.1), given the many choices available. The scientific method to inform the design of embodied interactive agents and to assess its outcomes is through *evaluation*, which collects, analyzes, and interprets information about how a given agent works for a particular group of people in a specific situation, possibly with the potential to generalize these results and transform them into guidelines for the design of embodied interactive agents (cf. Section 15.1). Ideally, evaluation studies assess agents as they interact with real users in realistic scenarios over a longer period of time.[244]

This chapter is organized as follows. First, motivations for doing evaluation research on and with embodied interactive agents are identified in Section 15.1. Second, Section 15.2 discusses factors of agents, users, and contexts that influence evaluation. Different types of evaluation research with respect to the time and level of analysis involved are compared in Section 15.3. Methodology relevant for evaluating embodied interactive agents, including research models, research questions, constructs, research strategies, and methods for data collection and analysis, is reviewed in Section 15.4. Guidelines for the evaluation of embodied interactive software agents are provided in Section 15.5. Finally, Section 15.6 summarizes the chapter.

[244] While this is the ideal, restrictions on time, money, availability of subjects, and other resources as well as the difficulty of controlling real-world situations have typically led researchers to conduct short-term evaluations of agents with a relatively small number of users in a controlled laboratory setting.

15.1 Motivations for Evaluation

The evaluation of embodied interactive software agents involves the goal-directed, methodical, and systematic collection, analysis, and interpretation of information about how a specific agent in a particular state of development works for a particular group of users (cf. Chapter 6.1) in a particular predefined context of use (cf. Chapter 6.2) (Christoph 2004: 68). Evaluation is a crucial component of both research and development of embodied interactive agents (Sanders & Scholtz 2000: 1). It collects information about an agent that helps to answer questions of practical and/or theoretical interest to one or several of the parties involved, which include the following ([INT 170]):

- *Designers*. Evaluation produces sound and verifiable evidence and guidelines (see below) that inform decisions in design, demonstrate the suitability of the agent for the client's needs, and provide ammunition for the marketing campaign.
- *Clients*. Evaluation provides information for the paying customers that helps them to make decisions about intermediate and finished versions of the agent. Successive evaluations of increasingly refined designs and implementations can be scheduled as project milestones.
- *End-users*. Evaluation can (and should) allow end-users to become (and feel) part of the development process by eliciting their ideas, needs, preferences, and views regarding a particular agent developed for them. User involvement should start early and continue throughout the process of developing an agent (cf. Chapter 7.13).
- *Researchers*. Evaluation helps researchers to test theories targeting embodied interactive agents as such or using them as a vehicle to investigate other topics, such as principles of face-to-face human-human communication.

In general, evaluation is instrumental in focusing research and development efforts; identifying progress toward milestones and end results; selecting the best from among a set of candidates, and assessing success of the finished system (Sanders & Scholtz 2000: 1). Table 30 shows the major categories of research on embodied interactive agents, their criteria for success, and how they go about assessing it. Another important motivation for evaluation is to produce *guidelines* (i.e. general rules, best practices, pieces of advice, etc.) for the design of (different aspects of) embodied interactive agents, which are intended to streamline the design process and to ensure the quality of its outcome by helping designers to make choices that eventually produce an agent which is appropriate for its designated role and context (Ruttkay et al. 2004: 44f.). For example, evaluation could be performed to shed light on the role of personification (cf. Chapter 9.1) in designing the appearance and (communicative) behavior of agents.

Design guidelines may originate from the experience of practitioners in the field, or they may be based on a generalization of findings from one or several evaluation studies. A substantial body of guidelines for embodied interactive agent design has already been published in the literature (e.g. Reeves & Nass 1998; Cassell et al. 2000a; Payr & Trappl 2004; Prendinger & Ishizuka 2004a; Isbister 2006), which provided valuable insights for the discussion of agent design aspects in the previous chapters. Still, much remains to be clarified, detailed, and discovered with respect to what agent designs are effective for what users in what contexts.

Table 30. A taxonomy of major categories of embodied interactive agent research. Adapted from Isbister and Doyle (2002, Table 1; 2004: 11, Table 1.1).

Category/ Research Focus	Criteria for Successful Agent Designs	Evaluation Techniques	
		Subjective	Objective
Believability (“Illusion of life”)	The agent conveys the impression of being ‘life-like’ or “larger than life” to the viewer or user.	User or expert assessment of the believability of the agent’s appearance, voice, personality, responses, etc.	Comparison of the user’s behavioral and physiological reactions to the agent with his or her responses to other humans (similarities/ differences; attribution of human dispositions and qualities to the agent)
Sociability (Human-like social interactions between agents and users)	Users’ social interaction with the agent is ‘natural’ and ‘intuitive.’	Determining users’ perceptions of the agent’s friendliness, helpfulness, social and communicative abilities, etc.; users’ evaluation of their overall experience of the interaction (ease, satisfaction, speed)	Measures of elicited social responses to the agent; changes in user behavior predicted by the social tactics used by the agent (increased influence of the agent on the user’s answers, users’ willingness to reciprocate to the agent, etc.).
Application Domains (Design for different applications and roles)	The agent achieves the goals of the application (learning, entertainment, changes in attitudes and behaviors, etc.).	Measures of user satisfaction with task achievement and interaction; role usefulness of the agent as perceived by domain experts and end users	Testing of behavioral outcomes (memory, performance on tasks, etc.) using existing benchmarks; need to separate the effects of the agent from those of the application
Agency and Computational Issues (Algorithms, systems, and architectures)	The agents, components, and techniques developed meet criteria including elegance, parsimony, optimality, correctness, good performance on (common) benchmarks, broad applicability of solutions, etc.	Peer review by other researchers in the same knowledge area based on relevant publications, code inspection, use in other systems and contexts, etc.	Quantitative analysis and evaluation using formal (mathematical) techniques (e.g. faces or gestures recognized correctly, number of steps in the best plan, predictive accuracy of machine learning techniques, etc.)
Production (Polished end product; professional agent development)	The final agent exhibits good consistent quality throughout, with the look and feel of a professional product. The development process went smoothly and efficiently.	End-user measures of satisfaction with the agent, its level of professional quality, and the smoothness of the experience. Confirmation that users read the agent as intended	Empirical comparison of production tactics with other practices, using measurable benchmarks (time, money, human and other resources)

Design guidelines typically map from ‘independent’ aspects of an agent that can be manipulated by developers or experimenters to ‘dependent’ aspects that change as the result of the manipulation (Ruttkay et al. 2004: 44). Broad categories of designable agent characteristics include embodiment, behavior, expertise, communicative abilities, emotion and personality, which have been discussed in Chapter 9 to Chapter 13. The design of these independent aspects influences various dependent dimensions of the user-agent interaction, including:

- Users’ acceptance of the agent;
- The actual and perceived efficiency and effectiveness of the agent and the system in which it is embedded;
- The agent’s impact on the observable behavior, subjective experience, and objective performance of users.

It should be noted that these are not isolated but rather mutual effects. For example, the degree of acceptance of a pedagogical agent influences how long and how often learners will interact with it, which, in turn, may increase effectiveness if the interaction with the agent contributes to learning, or decrease effectiveness if dealing with the agent distracts the learner from the task. The positive or negative effects of agent design configurations can be modulated by another group of aspects related to the user and the context (Krämer & Bente 2001), which are discussed in Section 15.2.2 and Section 15.2.3 below, respectively.

Ruttkay et al. (2004: 44f.) pointed out that design guidelines may apply to specific tasks or application contexts, or they may be of more general use (for instance, there is substantial evidence that voice is a central feature for making an agent seem person-like, cf. Chapter 10.1.1). Other, even higher-level guidelines express complex relationships between multiple design parameters. In general, it is important to state the *scope* of design guidelines, i.e. the application type and user group for which they are valid (p. 45). For example, the high-level guideline “The agent should be consistent, i.e. its appearance, attitudes, behaviors, and language should all convey the same general impression of the individual it is intended to portray” (cf. Chapter 7.5) has general scope, i.e. it applies across applications and user groups. In contrast, the scope of the guideline “The agent should portray a personality that is similar to the user’s personality” is limited to people who tend to be attracted by similar personalities in their human interaction partners (p. 45).

15.2 Factors Influencing Evaluation

Evaluation studies on embodied interactive software agents normally involve a particular agent interacting with a certain group of users in a specific context. Agents, users, and contexts are each complex entities that can vary along multiple dimensions. In an (experimental) evaluation study, the aspects of all three elements, i.e. design parameters of the agent, characteristics of the user, and contextual factors have to be carefully controlled in order to create an evaluation environment in which only the aspects relevant to the current research question (cf. Section 15.4.2) are manipulated while the others remain stable (and thus do not interfere with the results).

15.2.1 The Agent

Users' reaction to and interaction with a particular embodied interactive agent are influenced by the way it looks, behaves, and communicates. Agents may differ from one another in terms of a wide variety of design parameters, and finding the 'right' configuration of parameters for a particular group of users and a specific role or task can be difficult. Design parameters for embodied interactive agents can be categorized as follows:[245]

Embodiment (cf. Chapter 9)

- *Personification*. How much does the agent persona resemble a human being in appearance?
- *Appearance*. What are the attributes of the agent's physical form in terms of species, demographics, body features, and outfit?
- *Completeness*. Does the embodiment consist of just a head; a head and neck; a head, neck, arms, and torso; or a fully articulate body?
- *Individuality*. Does the embodiment reveal and express the agent's unique and distinctive personality?
- *Movement*. What kinds of movement (if any) can the embodiment perform?
- *Presence*. Is the agent persona always present on the screen, or does it only appear when the user requests or needs the agent's assistance?
- *Realism*. Does the embodiment use or observe:
    - Two-dimensional (2D) .. three-dimensional (3D) graphical models;
    - Realistic .. artistic representations;
    - Physiologically accurate models .. cartoon shapes;
    - Human limitations .. cartoon actions?
- *Stability*. Does the agent remain in the same body across time and situations, or does it change its embodiment?

Communication (cf. Chapter 10.1 and Chapter 11)

- *Language*. Can the agent understand and produce language? What languages does the agent support? Does it provide text and/or speech input and output? Is the agent's speech intelligible, natural-sounding, and expressive? Are messages generated and delivered dynamically, or do they come from canned resources? What degree of complexity in inputs can the agent handle? Does the agent integrate language understanding and generation with other input and output modes?
- *Face*. What information does the face of the agent persona convey? Are facial displays synchronized to speech, gesture, etc.?

---

[245] Ruttkay et al. (2004) proposed another catalogue of design aspects of embodied interactive agents, which consists of the major categories embodiment, mental capabilities, implementation aspects, and application context. Xiao (2006) developed a conceptual framework for describing the design space of embodied interactive agent systems, which defined three categories of factors influencing the usefulness of agent-based interfaces: features of the agent, features of the user, and features of the task. The set of agent attributes consists of representation, presence, initiative, competence and performance, verbal communication, non-verbal expression, and role of the agent.

- *Gaze*. Does the agent persona use gaze to disambiguate the verbal message, send social signals, establish a channel to obtain information, regulate dialogue flow, express emotions, indicate personality characteristics, establish social control, and reflect levels of intimacy?
- *Gesture*. In what ways does the agent persona use its hands and other body parts to create movements which emphasize, clarify, or amplify its verbal message?
- *Posture*. What positions of the body does the agent persona support, and how are they used to indicate involvement, liking, and relative status?
- *Locomotion*. How does the agent persona move from one place to another? Is this movement influenced by the agent's personality, the agent's emotional state, the agent's physical constitution, the state of the environment, and cultural specifics?
- *Dialogue*. What mechanisms does the agent employ to manage the (multimodal) dialogue with the user? This includes techniques for allocating the initiative (system, user, or mixed), modeling the dialogue, modeling the user, and handling errors in the interaction.

Expertise (cf. Chapter 12)

- *Knowledge acquisition*. What mechanisms does the agent provide to create, maintain, and extend its domain knowledge (knowledge engineering and/or machine learning)?
- *Knowledge representation*. What formalisms are used to represent the agent's domain knowledge? Do the chosen formalisms facilitate authoring, learning, inference, problem solving, etc.?
- *Problem solving*. What algorithms does the agent use to solve problems?
- *Student modeling*. What information does the pedagogical agent collect about the learner? How is it represented and used? Does the information collected improve the agent's ability to assist the learner?
- *Instructional expertise*. What repertoire of instructional interactions for assisting the learner is available to the pedagogical agent? How are these interactions designed? Do they work for the target context and learners?

Emotion and Personality (cf. Chapter 13)

- *Emotion*. What spectrum of emotional states can the agent interpret, communicate/simulate, and express?

  - *Emotion modeling*. Is an emotion model used at all? Does the chosen emotion model emphasize deliberately communicated emotion or the simulation of 'true' emotion? Does the emotion model support different intensities and mixtures of emotions?
  - *Emotion interpretation*. What forms of sentic modulation (e.g. facial expression, voice inflection, blood pressure, etc.) can the agent perceive and interpret? How accurate are its interpretations of the user's emotional state?
  - *Emotion expression*. What means of communicating emotion (facial expression, voice inflection, gesture, posture, and movement) does the agent use? Are the agent's emotional displays easily interpretable and convincing (believable)?

- *Personality*. What personality was chosen for the agent? Is the personality appropriate for the agent's role?

  - *Personality modeling*. What framework is used to model the agent's personality (The Big Five, NEO-PI-R, etc.)?

- *Personality interpretation*. How good is the agent at interpreting aspects of the user's personality?
- *Personality expression*. What aspects of the agent's appearance, behavior, and communication convey its personality (ideally, all)?

Implementation

- *Animation*. What approach is used to produce the animations of the agent? Is the quality (defined as frames per second) of the rendered animations acceptable? What is the cost (computationally and financially) of producing the animations? Are the animations available at different levels of detail (Ruttkay et al. 2004: 39)?
- *Architecture*. What architectural framework (if any) is the basis for the agent? Is the chosen architecture appropriate for the agent's role and task? How difficult is it to adapt the agent to a new domain, platform, etc.?
- *Authoring support*. What tools does the implementation provide for application developers, in particular for non-experts?
- *Components*. How are the agent's features implemented in terms of a set of components? What is the ratio of proprietary and third-party components? Are all components of the same quality? Are the components designed for reusability?
- *Performance*. How fast can the individual components and the agent as a whole process inputs and generate outputs? Does the implementation support smooth human-agent interaction, without perceptible glitches and delays? Is the resource consumption of the implementation within acceptable limits?
- *Platform*. What is the targeted technical platform of the implementation? Does the implementation effectively make use of the resources and observe or work around the constraints of the platform?
- *Requirements*. What hardware and software are necessary to run the implemented agent? What are the assumed operational conditions for the implementation (noise level, screen size, network bandwidth, real-time interaction, etc.) (Ruttkay et al. 2004: 40)?
- *Standards*. To what extent does the implementation rely on and conform to standards and de-facto standards, e.g. of the World Wide Web?

15.2.2 The User

As discussed in Chapter 6.1, a user is a person who makes use of a thing to achieve something, for themselves or for other individuals, alone or in cooperation or competition with other users, for work or for pleasure. Quite a number of these 'things' are human-made artifacts, such as hand tools, weapons, and computers. Computer users interact with these machines through user interfaces (cf. Chapter 1.2) to carry out certain activities. The user interface of an embodied interactive software agent centers on an animated character and the aspects of face-to-face conversations it supports. How users interact with, assess, and are influenced by a particular embodied interactive agent depends on the individual characteristics of each user (Ruttkay et al.

2004: 48f.). These characteristics have already been described in Chapter 6.1 and are therefore just listed below as a summary:[246]

- Gender;
- Age;
- Culture;
- Education;
- Social class;
- Occupation;
- Organization;
- Intrinsic limitations;
- Extrinsic limitations;
- Skills;
- Goals;
- Needs;
- Expectations;
- Experiences;
- Preferences.

It is generally not feasible to include all members of the target population in an evaluation study. Thus, a *sample* or subset of the population has to be chosen (Christoph 2004: 86f.). Creating a sample involves carefully deciding *which* and *how many* individuals to select for the study. If generalizations from the sample to the population being studied should be possible, the sample has to be *representative* of the population. A non-representative sample is said to be *biased*. Representative samples can be obtained through *random sampling*, i.e. each member of the population has an equal probability of being selected, and the selection of one member does not affect the chances of the others to be included in the sample. When representativeness is not an issue (as in case studies, cf. Section 15.4.4.3), other sampling methods can be used (p. 87; Verschuren & Doorewaard 1999):

- *Minimize differences* between participants by selecting individuals that are very similar to each other. This strategy is useful in exploratory studies to avoid distractions by a potentially large number of variations when formulating hypotheses.
- *Maximize differences* between participants on relevant dimensions by selecting individuals that differ as much as possible, but only with respect to the aspects of interest (in all others, they are as similar as possible). With this strategy, a broad overview can be obtained.
- Select participants *one at a time*, based on the results of the previous study.

The number of participants to be recruited for an evaluation study (i.e. the *sample size*) depends on the *variability* of the phenomenon being studied. Phenomena with a wider range of variation require a larger sample size than others showing less variation. In general, larger samples allow more precise estimates of the phenomenon under study (Christoph 2004: 87).

15.2.3 The Context

Embodied interactive agents always operate in some kind of environment and thus need to be evaluated with respect to the *context of use* (cf. Chapter 6.2) for which they are intended. A number of contextual factors (Johnson et al. 2004a) can influence the evaluation, including:

The target user group

- *Number* of users (single user vs. multiple users);
- *Stance* of users toward the environment (first- vs. third-person perspective, god/director, or passive observer).

[246] Xiao (2006) identified a similar set of user characteristics that are relevant for the evaluation of embodied interactive agents. The characteristics are: perceptual abilities, gender, age, knowledge and prior experience, intent, personality, and ethnicity and social background.

The agent and other agents

- *Number* of agents (single agent vs. multiple agents);
- *Stance* of agents toward the environment (resident agents vs. migrating agents).

The task (largely adapted from Xiao et al. 2002; Catrambone et al. 2004: 250–253)[247]

- *Complexity* (number and nesting of subtasks and basic actions and the computational effort required);
- *Cooperation* (with the user, other agents, and/or external applications);
- *Criteria for success* (e.g. quality of performance and/or outcome, timeliness, resource consumption, user satisfaction, etc.);
- *Familiarity* (Do the agent and/or the user operate in a familiar or unknown domain?);
- *Focus* of the agent (The agent might be directly involved in the user's primary task or help with secondary tasks);
- *Isolated* activity (e.g. games, web sales, etc.) vs. *integrated* activity (part of a curriculum, work activity, ongoing interaction, etc.);
- *Objectiveness* (objective (information gathering, problem solving) vs. opinion-based (giving advice or recommendations));
- *Resources* (e.g. applications, databases, internal knowledge and models, input/output devices, other agents, etc.);
- *Temporal arrangement* (attention (constant or regular vs. on-demand), duration, and frequency).

The application domain

- Education (cf. Chapter 4), electronic commerce (cf. Chapter 3.2.2), entertainment (cf. Chapter 3.2.4), health care, information retrieval (cf. Chapter 3.2.3), etc.

The physical environment

- Public vs. private space; distractive vs. quiet environments; mobile vs. stationary systems; safety-critical vs. safe work environments.

The agent's social role

- Advisor, companion, guide, opponent, personal assistant, salesperson, tutor, etc.

The agent's role in the application

- *Major* vs. *minor* (the degree to which the application depends on the use of an agent);
- *Essential* (e.g. actors in a virtual drama, opponents in computer games, and agents providing a unique type of information or service);
- *Supporting/supplementary* (e.g. facilitators, guides, motivators, etc.);
- *Optional* (the application can be used with or without the agent);

[247] Xiao's conceptual framework (Xiao 2006) likewise includes aspects of the task: domain, duration and frequency, importance and consequence, and difficulty and objectiveness.

- *Temporary/phased* (supporting novice users, then gradually withdrawing as users' proficiency increases).

The life cycle of the interaction

- A *single episodic* interaction (e.g. information kiosks);
- *Repeated episodic* interactions (e.g. guides in web sites);
- Over *extended episodes* (e.g. interactive movies);
- Over a *series of extended episodes* (e.g. virtual course instructors);
- *Unending/lifelong* (e.g. ambient agents in the user's work or living environment).

15.3 Types of Evaluation Research

Naively, it could be argued that evaluation can wait until the final stage of development when the agent has been implemented and tested and is therefore ready to face the judgment of users. However, at that point, it may be too late to include missing features or remedy fundamental flaws in the design. In fact, some kind of evaluation should be performed during *all* stages of agent development (Christoph 2004: 69) to elicit users' opinions, needs, and preferences with regard to different features, designs, prototypes, etc. (cf. Section 15.3.1). Evaluation that happens while the agent is being developed is called *formative evaluation* (cf. Section 15.3.2), whereas evaluation at the end of the development process is referred to as *summative evaluation* (cf. Section 15.3.3).

With regard to the *level* at which evaluation should be performed, it is certainly important to assess the agent as a whole in the context of its application, using representatives of the target group of users as subjects. But evaluation should likewise determine the effects of individual components of the agent on the user's interaction with the agent and his or her overall experience. System-level or *macro evaluation* is discussed in Section 15.3.4 while component-level or micro evaluation is the topic of Section 15.3.5.

15.3.1 Evaluation in the Software Life-Cycle

The creation of large software systems for realistic applications is a complex task involving many different human talents (cf. Chapter 7.12), a considerable amount of time, and many interdependent subtasks. To ensure goal-orientation, productivity, and quality, software development has to be organized as a structured process. The most commonly distinguished stages of software development are described below (Christoph 2004: 69ff.). At each of these stages, evaluation research can be performed to inform the work to be done at that stage.

- *Task analysis/functional analysis*. *Task analysis* systematically decomposes a task into its elements (subtasks, activities, people, and resources) and indicates their relationships. *Functional analysis* identifies and describes the necessary functions of the system. Evaluation research can use unconstrained data collection methods, such as informal observations (cf. Section 15.4.5.3) of and unstructured interviews (cf. Section 15.4.5.2) with representative users to elicit information about the task and the functionality of the system.
- *Requirements specification*. A software requirements specification captures the interactions between users and the software in terms of use cases, defines a set of functional requirements describing the internal mechanisms of the software necessary to satisfy the use cases, and

specifies a number of non-functional requirements on the software, including performance requirements, quality standards, design constraints, etc. ([INT 171]). Ways to collect this information include more formal, structured, and focused observation and interviewing methods, as well as the use of focus groups (cf. Section 15.4.5.2), which involve a small number of people (users, developers, experts, etc.) in a moderated discussion of the desiderata of the final system.

- *Conceptual design/formal design*. *Conceptual design* creates detailed specifications of the requirements of the system (i.e. what the system is supposed to do, what data it needs for its operation, what users have to know in order to handle the system, etc.). *Formal design* specifies how the conceptual design can be achieved. Focus groups are a useful tool for generating ideas about the features of the system. However, the ideas produced have to be both feasible and compatible with the task and requirements specifications.
- *Prototyping*. Prototyping creates different working models of the system to be built, which are refined (or discarded) depending on the results of the evaluation [INT 172]). Prototypes are often tested using the Wizard-of-Oz setup (cf. Section 15.3.2), in which potential users are exposed to a prototype of the system and missing functionality is simulated by a human ‘wizard’ ([INT 173]). Prototyping is useful for developing preliminary ideas about the effects of different design features on representative users in particular settings.
- *Implementation*. Implementation creates programs and data stores for a particular platform (cf. Chapter 14.3). At this stage, having a more or less functional system allows to conduct experiments (cf. Section 15.4.4.2) and case studies (cf. Section 15.4.4.3) that evaluate the system with respect to, for example, functionality, reliability, usability, efficiency, maintainability, and portability. The goal is to obtain information for future extensions, refinements, and updates.
- *Evaluation*. The goal-directed, methodical, and systematic collection, analysis, and interpretation of information about a specific software system in a particular stage of development is an activity that ideally occurs at all the stages discussed above. Each stage poses different requirements with respect to what has to be evaluated and how. While evaluation activities at earlier stages are more explorative in nature, informing decisions to be made in design, evaluation at later stages focuses more on the assessment of outcomes of the design process.

Software life cycle models are used to represent all the activities involved in software development and how they relate to one another. The classic *waterfall model* (Royce 1970) views the software development process as flowing through a cascade of four discrete stages (design, coding, testing, and maintenance), with later stages providing feedback for earlier stages ([INT 174]). Later models describe software development as an iterative, user-centered process. *Iterative* means that the software is developed incrementally over a series of iterations ([INT 175]). Each *iteration cycle* contains all the stages of the development process (requirements analysis, design, implementation, and testing) and produces a functional partial version of the final system, which grows from one cycle to the next ([INT 176]). The development process is *user-centered* in that it is driven by the people who will use the system, and their goals, needs, and preferences, rather than by technology alone (Preece et al. 2002: 285). One example is the *spiral model* (Boehm 1988), which describes an iterative and incremental software development process that makes repeated use of prototyping (and other means) and performs risk analysis at every stage in order to keep the risks involved in the development process to a minimum ([INT 177]). Another very influential model of iterative, user-centered software development, the *star life cycle model* (Hix & Hartson 1993) shown in

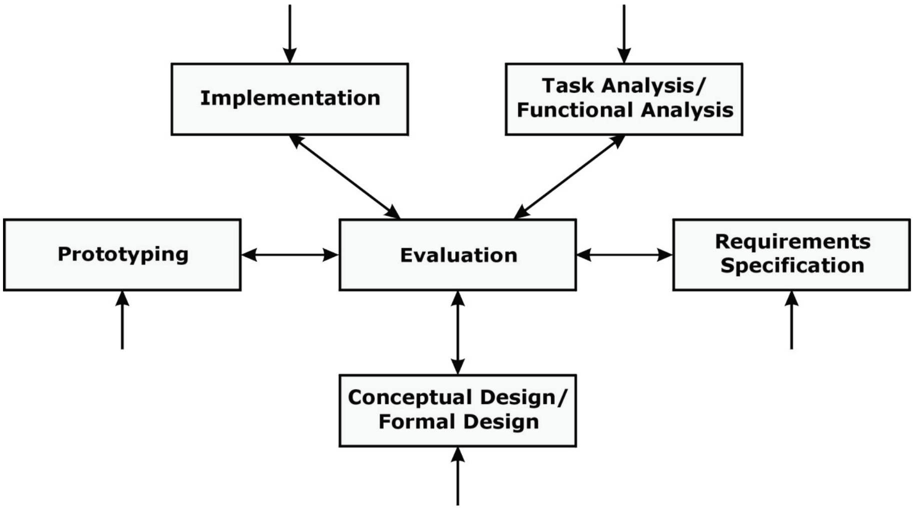

Figure 88. The star life cycle model. Adapted from Christoph (2004: 69, Figure 3.1).

Figure 88, explicitly acknowledges the central role of evaluation in the software life cycle by connecting all other stages to it (Stone et al. 2005: 19). In this model, the order of activities is less important than evaluation of all phases of development. As shown by the bidirectional arrows in the diagram, it is possible to move from the current activity to any other, provided that evaluation of its results happens first (Helms et al. 2006). Development can also start with *any* activity (indicated by the unidirectional arrows) (Costabile 2001; Stone et al. 2005: 19), rather than with an obligatory analysis of the task and the necessary functions, which reflects the common practice of entering the development process at different activities that is found in real-world projects. However, the high degree of flexibility provided by the star life cycle model may also be its greatest weakness because there is no indication where development should start or stop and, importantly, after how many iterations the process should end (Sevik 2003: 29). As a result, managing a project according to the philosophy of the star life cycle model can be a difficult task.

15.3.2 Formative Evaluation

> *When the cook tastes the soup, that's formative, when the guests taste the soup, that's summative. (Robert Stakes; quoted in Scriven 1991: 169).*

Formative evaluation is a *process-oriented*, typically exploratory type of evaluation, which collects information at various stages while the application is being developed, in order to shape ideas, guide decisions, and identify opportunities for improvement from early on (Sanders & Scholtz 2000: 2; Christoph 2004: 71f.). In general, formative evaluation studies should be cheap to perform because they recur regularly during the development process ([INT 170]). Important types of formative evaluation include:

- *Heuristic evaluation*. The goal of heuristic evaluation is to assess the usability (cf. Chapter 6.3) of a given user interface design. Experts carefully and systematically look over the user interface to identify problems related to the usability of its design, using *heuristics* that include established usability principles, such as visibility of system status, error prevention, aesthetic and minimalist design ([INT 178]), as well as their own knowledge of potential stumbling blocks for users (Nielsen 1994; Sanders & Scholtz 2000: 2). Heuristics establish themselves as valid rules of thumb for design through repeated successful application in the development of different systems, as determined by way of correlation with ratings of user satisfaction and quantitative measurements (Sanders & Scholtz 2000: 9).

- *Cognitive walkthrough*. This is another usability inspection method (Wharton et al. 1994) in which a group of evaluators walk through the steps of a task to be carried out via interaction between the user and the system and ask themselves whether the user will (1) realize that the current subtask is instrumental to achieving his or her goal and thus try to achieve it; (2) notice that the correct action is provided by the interface; (3) understand that the current subtask can be achieved by the action; (4) obtain feedback afterwards that he or she has performed the correct action ([INT 179]; [INT 180]). The focus is on identifying issues related to the comprehensibility and learnability of the user interface. The cognitive walkthrough method can produce results quickly at low cost, and it can be applied early in the design process, to a prototype, or even a paper mock-up of the interface ([INT 179]).
- *User studies*. This is the most important type of formative evaluation used in practice (Sanders & Scholtz 2000: 2). *Think-aloud studies* (Ericsson & Simon 1984) involve asking users to say aloud what they are thinking while they are working with a computer- or paper-based prototype of the system. The analysis of the resulting think-aloud protocols reveals what confuses them about the user interface (Sanders & Scholtz 2000: 2). *Wizard-of-Oz (WOZ)* studies[248] (Kelley 1984; Dahlbäck et al. 1993) involve real users interacting with a (partially) human-simulated version of the system. Subjects are led to believe that they are dealing with a machine, but the machine's responses are actually simulated by a hidden human 'wizard' ([INT 173]). There is no (fully) functional system behind the 'fake' conversational or other user interface (cf. Figure 89). Wizard-of-Oz setups have become quite popular for the prototyping and usability testing of user interfaces, including those with embodied agents (e.g. Buisine & Martin 2003; Chateau et al. 2005; Brown & Barrett 2006). The goal of these studies is to gain insights into how users would behave toward the system in a natural language dialogue (cf. Chapter 11.3) or some other kind of (unimodal or multimodal) interaction (cf. Chapter 11.2), before actually having the (complete) system in place.

In general, user studies and evaluator-driven approaches are complementary types of formative evaluation, as discussed by Sanders and Scholtz (2000: 8f.). Heuristic evaluation by experts can be used to systematically test a pre-defined set of abilities of a particular embodied interactive software agent. Evaluators can define *test cases* that cause the agent to exhibit certain qualities. In contrast, user studies cover only those abilities that were elicited by users during their interactions with the agent. However, these interactions may include unanticipated cases that generated unexpected responses from the agent. The information obtained from these unanticipated interactions is invaluable for assessing the robustness of the agent outside its familiar interaction patterns and for detecting opportunities for improvement.

15.3.3 Summative Evaluation

Summative evaluation is an *outcome-oriented* type of evaluation, which is performed toward the end of the development process, using the (almost) finished system (Sanders & Scholtz

[248] The name of this type of experimental dialogue scenario was chosen because such simulations recall the Wizard of Oz's deception in the 1939 movie ([INT 181]).

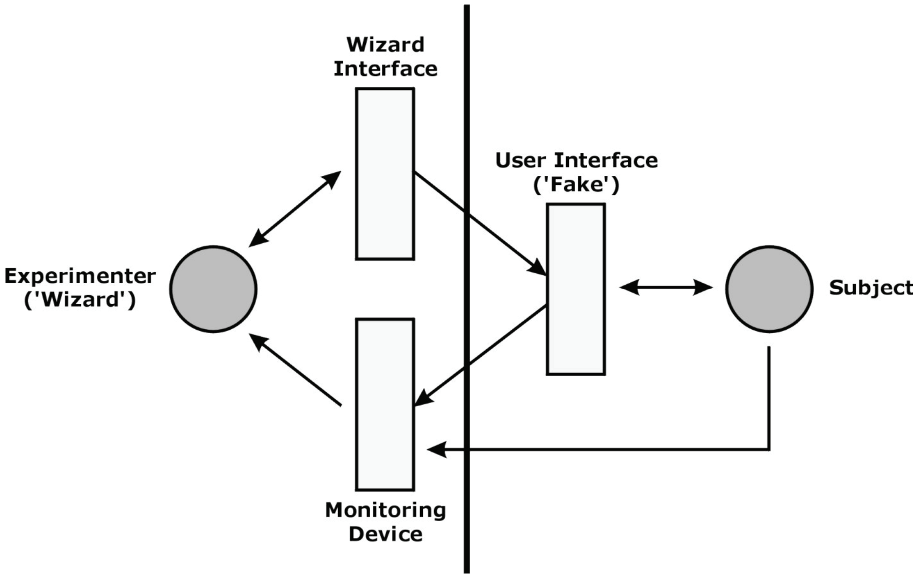

Figure 89. The basic setup of a Wizard-of-Oz experiment. Double-headed arrows mean 'interaction,' whereas single-headed arrows represent flows of information.

2000: 2; Christoph 2004: 71f.). The main goal of summative evaluation is to determine how users actually perceive and use the system and what can be improved in future versions. Therefore, a summative evaluation should show that people can (or cannot) use the system in its intended context or for its intended purpose ([INT 170]), and possibly also whether they like (or dislike) using it. In other words, summative evaluation has to test if the system is *accepted* by users who are representative of the target group ([INT 170]).

While formative evaluation (cf. Section 15.3.2) is largely exploratory, summative evaluation focuses on a well-defined set of research questions (cf. Section 15.4.2), which are answered on the basis of measures of performance (*benchmarks*) and user satisfaction (Sanders & Scholtz 2000: 1) collected in an experiment (cf. Section 15.4.4.2) or survey (cf. Section 15.4.4.1), using quantitative methods. Often, the results are *compared* to the measures obtained from other systems (of the same kind) that compete with the system being evaluated or are used as a baseline for comparison. Since users can differ considerably along multiple dimensions, (cf. Section 15.2.2), comparison studies need to control for differences between participants in order to exclude the possibility that observed differences can be attributed to the users taking part in the study rather than the qualities of the systems being compared. Each participant should work with all the candidates and spend the same amount of time with each of them. By measuring performance and user satisfaction, summative evaluation studies can help to identify the 'better' system. However, they do not reveal reasons for this superiority (Sanders & Scholtz 2000: 2).

15.3.4 Macro (System-Level) Evaluation

Macro (or system-level) evaluation investigates what a given embodied interactive software agent contributes to a particular application. The following questions are of interest in a macro-evaluation context (Ruttkay et al. 2004: 41f.):

- What kinds of value does a particular agent add to the application (cf. 7.1)? To answer this question, researchers have to evaluate the effects of a specific agent design in the context of the application and with a group of users that are representative of the target population.
- What agent is the most suitable for the application? Here, it is necessary to compare different configurations of agent design parameters (cf. Section 15.2.1) for the application and the target group of users.

Macro evaluation assesses the performance of the overall system consisting of the agent (cf. Section 15.2.1), the user (cf. Section 15.2.2), and the application, taking into consideration contextual factors (cf. Section 15.2.3) that might have an impact on the performance of this system. System-level evaluation has to be performed with a functional version of the agent that can interact with users, in a context of use that closely resembles the agent's target domain, and with a representative sample of real (potential) users of the system (Johnson et al. 2004a). It contributes to an understanding of the agent's role in the system by evaluating (Johnson 2004; Johnson et al. 2004a):

- *The user-agent interaction*. How efficient, fluent, pleasant, successful, etc. are the exchanges between the agent and the human user? Rather than just the agent itself, the evaluation addresses the ways in which the agent contributes to these aspects in the context of its interactions with users.
- *The user's experience*. What impression does the user get from its interaction with the agent? Unlike an observer, the user interacts with the agent in the context of a task or narrative and thus does not have a passive, observing role. Three major stages can be identified in the user's experience of an agent (or other system): the first encounter (the 'lure'), the first extended session, and repeated use over a longer period of time (Johnson 2004). Users may not pass through all these stages in their experience of an embodied interactive agent. In fact, only a few users so far have reached the third stage, and quite a number have never got beyond the first (e.g. users of the Microsoft Office Assistant, cf. Chapter 8.2). The user's experience is often assessed in the guise of his or her satisfaction with the agent and/or the user-agent interaction. *User satisfaction* is a multidimensional concept (Sanders & Scholtz 2000: 2; Ruttkay et al. 2004: 56f.), which consists of user perceptions of and responses to agent characteristics such as attractiveness (cf. Chapter 9.2), expertise (cf. Chapter 12), conversational abilities (cf. Chapter 11), voice quality (cf. Chapter 10.1.1), and personality (cf. Chapter 13.5).
- *The effectiveness of the agent-enabled application*. Is the agent instrumental in achieving the goals of the application in which it is embedded (e.g. promote learning, increase customer satisfaction, improve the user's health)? The application domain determines what measures of effectiveness should be selected for the evaluation. Usually, the effectiveness of an agent cannot be determined by measuring it once; instead, it has to be assessed over longer time periods. People may be drawn to an agent initially because of its novelty, but they may soon grow tired of it. Or they might be hesitant at the beginning but come to appreciate the agent eventually. In other words, while *short-term* evaluations can produce some suggestive evidence, it is usually problematic to use this evidence to predict the effectiveness of an agent in the long run (Kim 2005; Kim et al. 2007). However, *long-term* evaluation studies of embodied interactive agents (e.g. Bickmore 2003; Bickmore et al. 2005) with a timeframe of days, weeks, or months are still the exception rather than the rule. Furthermore, since embodied interactive agents are often included in applications to make these systems more accessible and to increase user engagement, a diverse sample of users should participate in the evaluation (cf. Section 15.2.2), especially if it is not possible to make reliable predictions

about the range of people that will use an agent, as in web-based applications. Finally, evaluations of effectiveness should not only look at the user's performance but should also examine the motivational effects of the agent, which are critical for learning, for example (cf. Chapter 12.3.8).

15.3.5 Micro Evaluation

Micro evaluation is concerned with the embodied interactive agent itself. Its goal is to assess how individual design parameters (cf. Section 15.2.1) of the agent or combinations of them impact the user's performance and/or his or her perception of the agent (Ruttkay et al. 2004: 41), preferably in a context of use (cf. Section 15.2.3) similar to the one of the overall system ([INT 182]). Important issues for micro evaluation are summarized below (Ruttkay et al. 2004: 41f.):

- Does an agent with specific attributes meet a set of expectations (of designers, clients, end-users, and researchers, cf. Section 15.1)? The object of the evaluation is a particular agent design, which is tested in a controlled setting. Such evaluations are important when design cannot resort to established guidelines (cf. Section 15.1) but has to rely on the designers' intuitions and skills. Since solid design guidelines for embodied interactive agents are still sparse, the latter is still a quite common strategy.
- What configuration of design parameters achieves desired attributes of an embodied agent (e.g. intelligible, natural-sounding and expressive speech, believable emotional expressions, capacity for delivering effective and timely instructional interactions, etc.)? To answer this question, it is necessary to test different configurations from the space of design options.

Embodied interactive software agents typically include implementations of relevant models of (human) audio-visual speech, non-verbal behavior, natural language, dialogue, personality, emotion, culture, etc. taken from the literature (on phonetics, conversation analysis, cognitive science, social psychology, anthropology, etc.) (Beskow et al. 2004; [INT 182]). A number of these models were discussed in the previous chapters. The *micro evaluation paradigm* (cf. Figure 90), as discussed by Beskow et al. (2004), is a framework for assessing if a particular model, in the way it has been implemented in a given agent, is understood by users interacting with the agent as intended by the agent's designers (cf. the relationship between the design model, system image, and user's model discussed in Chapter 6.3). This component-level evaluation is a necessary pre-condition for demonstrating the added value (cf. Chapter 7.1) of embodied interactive agents at the system level (macro evaluation). *Judgment studies* involve implementing a given model in an agent and testing with users if the agent appears and behaves in a way that is close to the model ([INT 182]). Usually, the desired information cannot be reliably obtained by asking participants *directly* whether the agent being evaluated possesses certain properties, such as competence, likeability, and trustworthiness. Instead, this should be tested *indirectly* by engaging participants in an interaction with the agent showing that they assign these properties to the agent ([INT 182]).

Many models borrowed from the literature are incomplete, controversial, and were not designed to be implemented in embodied interactive agents, lacking necessary information or detail (Beskow et al. 2004). This is especially true of higher-level models (e.g. of emotion and personality, cf. Chapter 13), which describe aspects of an individual that are not directly observable, in contrast to lower-level models (e.g. of audio-visual speech and non-verbal behavior), which account for aspects that are more open to inspection ([INT 182]). To fill gaps

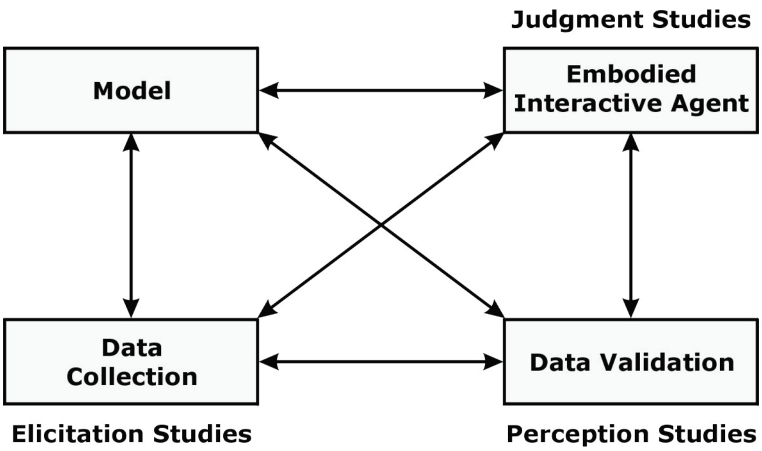

Figure 90. The micro evaluation paradigm for embodied interactive agents. Adapted from Beskow et al. (2004).

in a particular model or obtain information for building data-driven models for embodied interactive agents, it is necessary to collect data (from human subjects) and check its validity. This is done in *data elicitation studies* and *data validation studies*, respectively (Beskow et al. 2004; [INT 182]).

15.4 Evaluation Methodology

Embodied interactive software agents have to be evaluated in a methodologically sound way. Fortunately, the design of agent evaluation studies can draw on a vast body of existing know-how for doing evaluation research. The following sections review different aspects of this know-how in the context of evaluating embodied interactive agents. The discussion is largely based on two articles by Christoph (2004) and Ruttkay et al. (2004) that were published in the volume on evaluating embodied interactive agents[249] edited by Ruttkay and Pelachaud (2004).

15.4.1 The Research Model

Any evaluation study should be based on a clear view of the basic research design. The research model provides a visual, schematic overview of a particular evaluation study that shows the main stages involved in the research, which include the following (Christoph 2004: 72ff.):

- *Formulation of the research objectives.* Researchers may intend their evaluation study to contribute to the building or extension of a theoretical framework, or they may hope that it helps them (or others) to solve some practical problem. A third possibility is that they may be interested in deriving design guidelines for embodied interactive agents (cf. Section 15.1).
- *Description of the nature and setup of the study*, i.e. what kind of research will be conducted, and how. For example, *theory-testing* research is concerned with the formulation and testing of *hypotheses* (i.e. tentative and testable suppositions or explanations to be confirmed or

[249] In the contributions to that volume, these agents are called "embodied *conversational* agents."

disproved (*falsified*) by further investigation[250]). Evaluation research with a *diagnostic* perspective attempts to determine the nature or cause of a particular phenomenon.

- *Identification of information sources.* This typically involves reviewing relevant publications from the research literature (cf. Section 15.4.4.4) and the interviewing of domain experts (cf. Section 15.4.5.2).

An example of a research model for evaluating an embodied interactive agent is shown in Figure 91. Such a research model provides the basis for formulating the main question to be answered by the evaluation research.

15.4.2 The Research Question

The research question identifies the phenomenon that researchers plan to address in a particular evaluation study in order to achieve their research objective ([INT 185]). A strong research question clearly indicates what has to be done and what data has to be collected in the study, thus guiding the research and making it more efficient (Christoph 2004: 74). Good research questions provide information about the domain of the study, the constructs and variables to be used, the target population, the time and place (if they are relevant for understanding the study and/or interpreting its results), and the type of problem (e.g. descriptive, explanatory, or design-related) with which the study will be concerned (pp. 74f.). Furthermore, the research question should be explicit about the anticipated benefits of answering the research question and the range of potential benefactors (and how the results of the research can help them) ([INT 185]). Research questions with a narrow focus can more feasibly be answered by a single evaluation study than broad questions. For example, the question “How can designers make agents more believable?” is too broad to be answered by a single evaluation study (and there may not even be a single, universally applicable answer). A more specific question would be a better starting point, such as: “What is the relationship between facial displays of anger, joy, and sadness and the perceived believability of an embodied interactive agent?” ([INT 185]).

15.4.3 Constructs

After the research question has been formulated, it is necessary to find a way to describe and measure the phenomenon of interest. Often, evaluation research addresses abstract phenomena which cannot be measured directly, such as the user’s subjective experience of an embodied interactive agent, the agent’s influence on the user’s attitudes and behaviors during and across interactions, and effects of the agent on the outcome of the interaction. These phenomena are described in terms of *constructs* (Christoph 2004: 75f.). A construct, or *psychological construct* as it is also called, is a hypothesized latent entity which is an abstract characterization of observable human behavior. It variably assumes the name of ability, attribute, trait, and other mental states, characteristics, or qualities that cannot be measured or observed directly (p. 75). Examples of constructs that may be involved in the evaluation of embodied interactive agents include trust, liking, frustration, motivation, intelligence, language proficiency, and personality.

---

[250] A hypothesis can only be confirmed or disproved if it is *falsifiable*, i.e. if it can be shown to be false by way of performing an experiment or making an observation ([INT 183]). However, even if confirmed, the hypothesis is not necessarily proven but retains its provisional nature ([INT 184]).

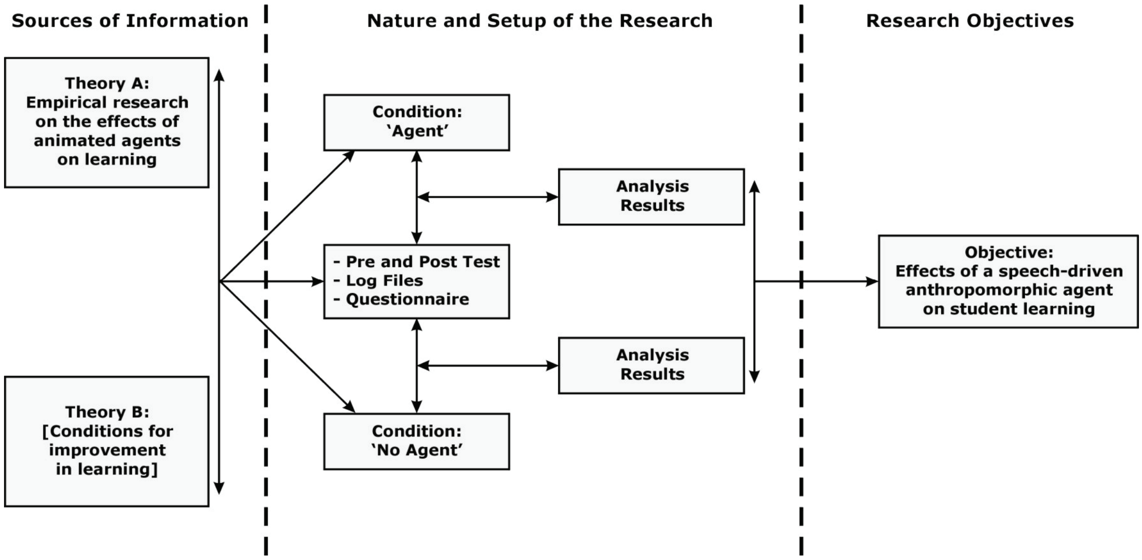

Figure 91. A research model for an empirical evaluation study aiming to determine effects of a speech-driven anthropomorphic agent on student learning. Adapted from Christoph (2004: 73, Figure 3.2).

The intangible nature of constructs requires their *operationalization*, which involves the translation of a given construct into a set of variables[251] for which researchers can obtain values directly through measurement or observation (Christoph 2004: 75, cf. Figure 92). For example, anger can be operationalized by measuring or observing facial expression, word selection, loudness and tone of voice ([INT 186]). For subjective constructs like trust and liking, finding an appropriate decomposition into measurable variables that can be related to aspects of agent design can be difficult a priori since these notions depend on the application context and on the user (Ruttkay et al. 2004: 47). In general, *construct validity* has to be verified in order to make sure that the instrument chosen measures the latent construct for which it has been adopted (Christoph 2004: 75; [INT 187]).

Ryu and Baylor (2005) developed and validated an empirical instrument for measuring learners' perception of a pedagogical agent persona. Their final agent persona instrument defined two constructs: *Informational Usefulness*, which corresponds to human instructor ability and refers to the pedagogical agent being perceived as a knowledgeable instructor; and *Affective Interaction*, which corresponds to human instructor personality and communication and refers to the agent's caring and motivating social presence (cf. Chapter 5.4.5) and its ability to communicate naturally with the learner, including emotional expression (cf. Chapter 13.4.2) and non-verbal communication (cf. Chapter 10.1) (Ryu & Baylor 2005: 307ff.). Four factors of a pedagogical agent persona emerged: the first two, *Credible* and *Facilitating Learning*, constituted the construct of Informational Usefulness, whereas the remaining two factors, *Human-like* and *Engaging*, made up the construct of Affective Interaction. For each of the four factors, Ryu and Baylor proposed a set of questionnaire items (Ryu & Baylor 2005: 315), with each item measured in terms of a 5-point Likert scale (cf. Section 15.4.5.1).

[251] Variables are measurable characteristics of an individual or system that can assume different *values* and have a hypothesized influence on the situation being studied. A *testable* variable can capture at least the presence or absence of a particular characteristic, i.e. it has two (or more) values (Buisine et al. 2002). Variables can be aggregated and measured at different levels (Christoph 2004: 75).

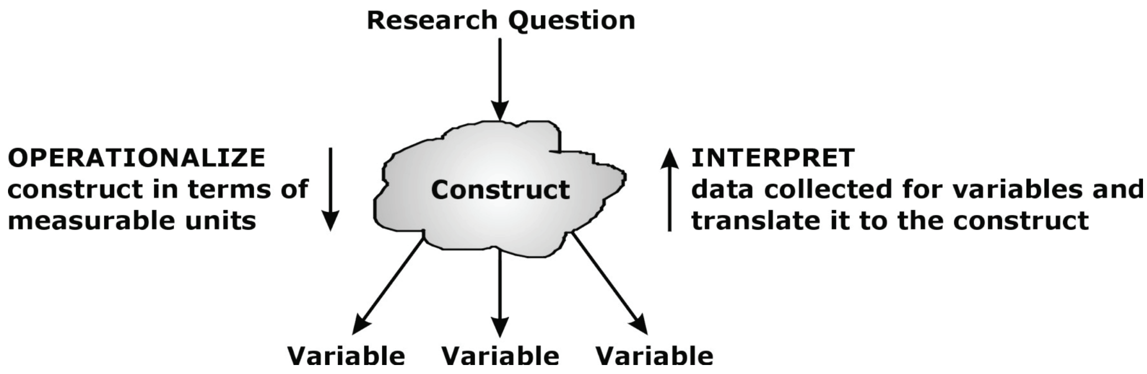

Figure 92. The operationalization of a construct into variables. Adapted from Christoph (2004: 76, Figure 3.3).

15.4.4 Research Strategies

> *A research strategy is like an airport or a bus terminal. The terminal is not the destination. It is a facility that helps travelers get where they want to go. (James E. Page).*

Once researchers have a model for their research, a solid and feasible research question, and constructs and variables for measuring phenomena of interest, the next step involves devising a *research strategy*, i.e. a plan for conducting the evaluation research in a systematic and goal-oriented manner (Christoph 2004: 76f.). Depending on whether researchers choose to collect their own data or to review the data collected by others, two major categories of research strategies can be identified. The first, *empirical research* (Beller 2004; Höfer & Tichy 2006), conducts systematic and controlled inquiry into a domain of interest that involves the first-hand collection of data from experience, observation, or experimentation, followed by the analysis of the collected data according to established and well-defined principles ([INT 188]). Three empirical research strategies are important for the evaluation of embodied interactive agents (Christoph 2004: 77):

- *Survey* (cf. Section 15.4.4.1) and *experiment* (cf. Section 15.4.4.2) for summative evaluation (cf. Section 15.3.3);
- *Case study* (cf. Section 15.4.4.3) for formative evaluation (cf. Section 15.3.2).

In contrast, *desk research* (cf. Section 15.4.4.4) is the process of reviewing available material (data and publications) to answer a new research question (Christoph 2004: 77). All four research strategies (survey, experiment, case study, and desk research) are discussed in the following subsections.

15.4.4.1 Survey

Surveys (Krosnick 1999; Timmerman 2002; Glasow 2005; Dillman 2007) involve the systematic collection of data from a typically large number of individuals (more than 50), which are randomly selected rather than chosen deliberately (Christoph 2004: 77), by asking them individually a series of pre-determined questions (Houston 1997). Popular techniques for conducting surveys include low-cost and low-effort methods, such as questionnaires consisting of mainly closed questions (cf. Section 15.4.5.1) in paper or electronic format, telephone surveys, and structured interviews (cf. Section 15.4.5.2) (Christoph 2004: 77). To protect the privacy of participants, surveys are typically *anonymous* or *anonymized* (Leone et al. 2006: 20).

Surveys provide a flexible and efficient way for developers of embodied interactive agents to collect a variety of information about the target group of people that will be using their agents (e.g. users' attitudes, background, experience, wishes, etc.) from a relatively large subset of users ([INT 189]). A thoughtful survey design combined with a high number of respondents allows the generalization of the survey results to the overall population. Often, surveys produce quantitative (numerical) data (cf. Section 15.4.5) that can be analyzed statistically (cf. Section 15.4.6.2) to investigate the relationships among (groups of) variables (Leone et al. 2006: 20). Well-defined techniques are available to assess if the results of a survey are valid, reliable, and statistically significant ([INT 189]). Surveys can also be used to compare subgroups of the population or results from one survey to the next, and to collect data that is representative of or allows predictions about the target population (Leone et al. 2006: 19). By using a standardized set consisting only of questions pertaining to the research, surveys provide for economy of data collection and analysis ([INT 189]). Finally, the data collected from electronic surveys can be automatically stored and pre-processed, which facilitates both data collection and analysis.

However, these advantages should be weighed against several shortcomings. For one, surveys will only generate useful data if they have been carefully designed. But even well-designed surveys can fail to reveal important (especially unanticipated) aspects if they only consist of pre-determined questions and answers (Leone et al. 2006: 20). Surveys adopt the individual as their unit of analysis, which may not be the best choice for studying social processes and other complex phenomena ([INT 189]). In general, surveys have to rely on participants' motivation, honesty, and memory and that they have the time, technology, and mental/physical ability to respond. But people may not know or remember, they may be sloppy or deceitful, and they may wish to project a positive image of themselves ([INT 189]). They may also be inclined to give answers which they assume are 'expected' of them because those answers are "politically correct" or suggested by the format or content of survey items. Paulhus defined this *response bias* in general as "a systematic tendency to respond to a range of [survey] items one some basis other than the specific item content (i.e. what the items were designed to measure)" (Paulhus 1991: 17). Response bias affects survey validity (Leone et al. 2006: 19). People who choose to respond to a survey typically select themselves into a group with certain properties. For example, only people with a positive attitude toward computer technology may choose to fill in a computer-based survey. This *self-selection* can bias the results of a survey because of the potentially large number of non-respondents (Christoph 2004: 77). Therefore, surveys require high response rates to produce reliable results that are generalizable (Leone et al. 2006: 19f.). Both the response bias and the response rate are difficult to control in a survey study (p. 20).

15.4.4.2 Experiment

An experiment (Campbell & Stanley 1963; Weber & Skillings 2000; Heffner 2004a; Christensen 2006; Martin 2007; [INT 190]; [INT 191]) is an empirical scientific procedure performed to discover new things, support or falsify hypotheses (cf. Section 15.4.1), or demonstrate what is already known (OED 1998: 647; [INT 191]). Experiments are used to study *cause and effect* in a controlled setting by deliberately manipulating one variable and keeping all other aspects of the situation constant ([INT 190]).

Participants are randomly assigned to one of two groups (conditions), the *experimental group* and the *control group*. A change (the *intervention*) is introduced for the experimental group but not for the control group. The control condition is the standard or *baseline* to which the results of the experimental condition are compared ([INT 190]). Nass et al. argued that the comparison of *varied conditions* is necessary if the hypotheses to be tested in the experiment are relative and comparative in nature (e.g. "Users will perceive Agent A as more instructor-like than Agent B" or "Users will find it easier to remember the agent's messages if they are delivered using speech rather than text"), which they typically are in experimental studies on embodied interactive software agents (Nass et al. 2000a: 376).

It is essential to *match* the experimental and control groups on all relevant characteristics, including age, gender, experience, etc. ([INT 190]). According to Nass et al., if the differences between the experimental and the control condition are restricted to one or two well-defined characteristics (e.g. use of synthetic speech vs. recorded speech), the results allow very specific conclusions that provide answers to specific design-related or theoretical questions. However, larger effects will likely go unnoticed. In contrast, if the conditions differ at a gross level (e.g. agent vs. no-agent), only very broad conclusions are possible, which may be useful at the early design stages (e.g. should the application feature an agent at all?) but do not help with particular design decisions (Nass et al. 2000a: 376).

Nass et al. provided two reasons why *random assignment* of subjects to the conditions of an experiment is important. First, if users are deliberately assigned to a particular condition based on gender, computer literacy, personality, or some other characteristic, observed differences between conditions cannot be attributed clearly to the manipulation but may also result from the properties or qualities of the subjects in a condition. Second, if users can choose their condition, the result may be an imbalanced distribution of subjects among the conditions. Furthermore, outcomes of the experiment may be influenced by subjects liking either the condition they chose (*post-decision justification*) or the other condition(s) which they did not choose (*buyer's remorse*) (Nass et al. 2000a: 376).

Experiments can be used to systematically evaluate the effects of design parameters of embodied interactive agents (cf. Section 15.2.1) by varying the value of a single parameter at a time while keeping the rest constant (Ruttkay et al. 2002). However, generalizations from the results of a single experiment may be problematic. But it is possible to *replicate* experiments, and the confidence in the validity of generalizations from an experiment will increase as more and more researchers repeat the experiment with the same results ([INT 190]). Replication presupposes that the research question (cf. Section 15.4.2), the constructs and the variables operationalizing them (cf. Section 15.4.3), the agent being evaluated (cf. Section 15.2.1), the group of subjects (cf. Section 15.2.2), the context (cf. Section 15.2.3), the experimental setup, and the methods for data collection (cf. Section 15.4.5) and data analysis (cf. Section 15.4.6) are clearly described. Well-established metrics and methods should be preferred to facilitate replication of the experiment by other researchers.

15.4.4.3 Case Study

Case studies (Stake 1995; Yin 2003) involve the in-depth, long-term study of the development of a particular individual, group of people, event, or situation (the *case*) (OED 1998: 282; [INT 192]). The case is described in terms of rich, mostly qualitative data (cf. Section 15.4.5), which helps the researcher to develop a comprehensive understanding of the case. This, in turn, is instrumental in identifying directions for future research. In particular, case studies can help researchers to formulate hypotheses, which can be tested by follow-up experiments (cf. Section 15.4.4.2) ([INT 192]).

Because of their detail-oriented, long-term nature, case studies normally involve only a small number of participants (typically at most 8 to 10), which are selected deliberately rather than randomly (unlike in experimental studies). In contrast to surveys (cf. Section 15.4.4.1), case studies collect data using more time-consuming methods, such as interviews (cf. Section 15.4.5.2) and observations (cf. Section 15.4.5.3). The data obtained is subjected to a qualitative rather than a quantitative analysis (cf. Section 15.4.6.1). All in all, case studies are a very flexible research instrument, but also one that (normally) does not produce generalizable results and may require a lot of effort (Christoph 2004: 79).

Kroetz (1999) conducted a case study on a virtual reality training simulation featuring the pedagogical tutor agent STEVE (cf. Figure 17b). To shed light on the effects of adding an animated tutor to the simulation, three constructs (cf. Section 15.4.3) were adopted: tutor believability, active participation by the participants, and the mental models that they form of the simulation (cf. Chapter 6.3). Data was collected from relevant publications, personal interviews, system documentation, the program environment, and a simulation procedure, in which 10 deliberately selected individuals were studied as they interacted with STEVE, observing an operation procedure demonstrated by the agent and then performing the procedure themselves. The results of the study indicated that STEVE was believable as a sentient virtual human (although it did not express emotions), induced participants to respond to the agent in human-like ways, helped them to efficiently learn a complex procedure, and ensured that they actively participated in the simulation (Kroetz 1999).

15.4.4.4 Desk Research

While empirical research is concerned with collecting and analyzing *primary data*, i.e. 'raw' data obtained directly from the source, *desk research* finds and analyzes *secondary data*, i.e. existing data which has been collected, processed, and published before, typically by other researchers and for a different purpose. Sources of secondary data include books, journal articles, conference proceedings, web sites, discussion forums, newsgroups, mailing lists, weblogs, corpora, databases, statistics, and others.

Often, desk research is conducted as a preparation for other kinds of (empirical) research, to find out about the work of others, identify interesting research questions (cf. Section 15.4.2), inform study design, etc. Successful desk research involves ([INT 193]):

- *Knowing where to look and what to look for*. Desk research should start with compiling a list of sources that provide both information and pointers to other sources.
- *Understanding the quality of the source material*. Sources differ with respect to how they obtain, analyze, and process their information, and for what purposes. Hence, the challenge is to harmonize information from different sources and to identify the best sources for the study.
- *Obtaining the right information*. Most information might be interesting, but it is not important to answering the research question.

Meta studies are a particular type of desk research which reviews secondary data published in other studies and draws conclusions from their results, methods, etc. (Christoph 2004: 77). Previous meta studies on embodied interactive software agents have identified, for example:

- The impact of animated interface agents on the user's subjective experience of the overall system, his or her behavior during interactions with the system, and the outcome of the interaction measured in terms of the user's performance (Dehn & van Mulken 2000);

Table 31. Data collection methods for empirical research strategies. Adapted from Christoph (2004: 80, Table 3.2).

Data Collection Method	Survey	Experiment	Case Study
Questionnaire	Yes	Yes	
Interview	Yes		Yes
Observation		Yes	Yes
Log Files		Yes	
Physiological Measures		Yes	

- Major categories of embodied interactive agent research (Isbister & Doyle 2002);
- Criteria for agent quality (Haddad & Klobas 2002);
- A framework for summarizing evaluation studies on embodied interactive agents (Ruttkay et al. 2002);
- A conceptual framework for describing the design space of embodied interactive agent systems (Xiao 2006).

15.4.5 Data Collection

Data collection is the process of gathering information from the participants in an evaluation study. What method is chosen to collect the data for the study depends largely on the research strategy adopted (Christoph 2004: 80). Table 31 lists the methods commonly used for the different empirical research strategies discussed in the previous section. These methods fall into two categories:

- *Qualitative methods* collect detailed *descriptive, non-numerical data* in audio, visual, or verbal format that is captured on a *nominal* scale[252]. Common examples of qualitative data include interview transcripts, field notes, video and audio data, images, and documents (e.g. reports, minutes, e-mails, stories, etc.) but also colors, smells, tastes, textures, brands, ethnic backgrounds, marital status, feelings, judgments, and so on.
- *Quantitative methods* collect data that is counted or measured (on an *ordinal*,[253] *interval*,[254] or *ratio* scale[255]) and expressed as numbers. Many kinds of survey data (cf. Section 15.4.4.1) are of this type, in particular Likert scale category responses (cf. Section 15.4.5.1). Other examples include test scores, checklists, time to complete a task, age, weight, heart rate, IQ scores, voltage, temperature, etc. Nominal data is also sometimes regarded as quantitative data.

---

252 Nominal scales provide a set of distinct, unordered categories that can be attached as labels to items ([INT 194]). Examples of nominal categories include male/female (participants), positive/negative (judgments), different colors, etc. (Christoph 2004: 80).

253 Ordinal scales measure degree (but not amount) of difference in terms of an ordered set of categories ([INT 195]).

254 Interval scales consist of ordered categories with an equal distance between them ([INT 196]). Their zero point is arbitrary ([INT 194]).

255 A ratio scale is characterized by ordered, equidistant categories, and it has an absolute zero value ([INT 197]; Christoph 2004: 80).

Data collection methods should generate *reliable* data, which is *repeatable* or *reproducible*. That is, similar results should be obtained if the study is performed again with the same method under the same circumstances (Christoph 2004: 88). The data collected should also be *valid*, i.e. accurately reflect the construct being measured (pp. 90ff., cf. Section 15.4.3). The most informative and valid data can be obtained from using a combination of data collection methods in the study (e.g. questionnaire, observation, and interview) (p. 80).

15.4.5.1 Questionnaire

Questionnaires (Houston 1997; Krosnick 1999) are by far the most popular data collection method used in conducting surveys (cf. Section 15.4.4.1), but they are also used to collect data in experiments (cf. Section 15.4.4.2), for example to quantify the effects of the personification of agents (cf. Chapter 9.1) on the user (e.g. Koda & Maes 1996; Lester et al. 1997b; Cassell & Thórisson 1999). A questionnaire consists of a set of questions, on paper or in electronic form, which participants are invited (by mail, in person, or over the Internet) to answer. The filled-in questionnaires are collected and analyzed. Questionnaires may involve two kinds of questions (Christoph 2004: 82f.):

- *Open-ended questions* give participants the opportunity to freely formulate their answers. The resulting qualitative data has to be interpreted manually, which requires more time and can lead to less objective analyses. In addition, it may be difficult to motivate participants to write longer answers.
- *Closed questions* provide participants with a limited set of answering options to choose from. There are several different types of closed questions, including semantic differentials, rating scales (in particular the *Likert scale* for measuring participants' degree of agreement or disagreement with a statement on a scale with an odd or even number of categories that allows them to give a neutral answer or forces a choice, respectively), item ranking, and option selection (Houston 1997). Closed questions generate quantitative data.

15.4.5.2 Interview

The *research interview* is a more or less structured one-to-one social interaction between a researcher and a subject. The interviewer attempts to collect data from the interviewee by asking him or her questions, face-to-face, over the telephone, or via the Internet, that aim to elicit the interviewee's views, knowledge, or whatever aspect is the focus of the research.

Structured interviews are formal, systematic, and standardized exchanges in which the interviewer asks each participant the same catalogue of pre-determined questions in the same order and wording in order to maximize consistency from one interview/interviewee to the next (TV&C 2001; [INT 198]). For the most part, structured interviews use closed questions, and they generate quantitative data for statistical analysis. Structured interviews are appropriate in situations where researchers are (quite) certain about what data they wish to collect and the data can be expressed in quantitative terms ([INT 199]). The range of replies should be predictable, whereas the magnitude of each reply has to be determined (Bonato et al. 2007: 38).

Unstructured interviews are spontaneous, free-flowing conversations between interviewer and interviewee ([INT 199]), which follow some general themes but are mainly intended to give respondents freedom in talking about the topic and to allow them to steer the interview in a different direction. The data obtained during an unstructured interview is normally qualitative in nature. Due to their flexible nature, unstructured interviews have applications in explorative studies where the nature of the desired information is not known ([INT 199]), for example to develop a general understanding of the problem to be solved at the beginning of a project. For instance, unstructured interviews may be conducted with representative users from the target group of an embodied interactive agent to be designed in order to find out about their attitudes toward and previous experiences with agents and other kinds of intelligent user interfaces. However, researchers require considerable interview skills to carry out successful unstructured interviews because they have to maintain the overall focus of an interview while giving the interviewee as much freedom as possible.

Semi-structured interviews are somewhere between the two previous types. On the one hand, the researcher follows a pre-determined schedule for the interview specifying the questions and their order, but on the other hand, the interviewer may deviate from the schedule to explore interesting topics or issues that emerge during the interview ([INT 199]). This hybrid interview form is particularly useful in studies where researchers have a fairly good idea of the questions or issues of interest but are less sure about the range of responses from the participants (Bonato et al. 2007: 38). In general, the need to integrate flexible exploration with the pre-specified schedule presupposes quite sophisticated interview skills on the part of the person conducting the interview.

Focus groups (Powell & Single 1996; Gibbs 1997; Morgan 1997) are small groups of people (up to 10) that are brought together by researchers to discuss and share their views on the topic under investigation in an interactive group setting (Powell & Single 1996: 499). The group interaction is guided and facilitated by a *moderator* who is responsible for ([INT 199]):

- Making transparent the purpose and goals of the group;
- Putting participants at their ease;
- Facilitating interaction within the group by asking open questions;
- Requesting clarification, details, or examples from the participants;
- Maintaining progression and focus of the discussion;
- Making sure that each group member contributes to the discussion.

Focus groups discussing the design of an interactive software system (e.g. an embodied interactive agent) are commonly shown a computer- or paper-based prototype of the system and asked to express their attitudes, feelings, impressions, etc. with respect to the design prototype. It is essential that the members of the focus group can do this freely. Therefore, the focus group should not be moderated by a developer of the system under review because he or she would be unable to perform the duties of the moderator with the necessary neutrality (Christoph 2004: 81).

15.4.5.3 Observation

Observation is a process that involves the carefully planned, prepared, and executed perception of a person or thing in a specific environment over a certain period of time in order to obtain information. “An observation is always preceded by a particular interest, a question, or a problem – in short, by something theoretical” (Popper 1979: 342).

Formal or *systematic* observation works with a simplified view of reality that is created by selecting and studying particular aspects of this reality. This complexity reduction is achieved by following explicit rules in order to facilitate replication of the results by other researchers. The observation produces quantitative results indicating, for example, the frequency with which a particular observed behavior occurs (Christoph 2004: 83f.). Systematic observation involves the two principles of *selectivity of perception* (i.e. the need to choose a subset of reality for observation) and *subjectivity of perception* (i.e. the tendency of observers to make mistakes and to be biased by their previous experiences, knowledge, and ideas) (p. 84).

Informal or *descriptive* observation is an unstructured and exploratory process that is usually carried out when little is known about the target population and its behavior. The researcher observes the behavior, conversations, and other interactions of the participants. This kind of observation generates qualitative data in the form of notes, sketches, pictures, audio and video recordings, and so on (Christoph 2004: 81). For example, the researcher can observe a master teacher of linguistic fieldwork who supervises a group of apprentice fieldworkers as they interact with one or several native speakers to elicit linguistic data, and take notes of any aspects of the roles and behaviors of the native speaker, the master teacher, and the apprentices that could inform the design of embodied interactive software agents for the roles of virtual native speaker, coach, and peer (cf. Chapter 16.2). Informal observation helps to refine research questions (cf. Section 15.4.2) and develop appropriate research strategies (cf. Section 15.4.4) when a research project is still in its early stages (p. 81).

Participant observation (Spradley 1980; DeWalt & DeWalt 2002) is a kind of informal observation where the observer is directly involved with the people, environment, and events that he or she is studying ([INT 200]). The researcher's immersion in the group of participants allows him or her to become privy to events which participants might be reluctant to share with an outside observer. However, a participant observer will never be objective, and keeping participants in the dark about the fact that they are being observed by one of their own group gives rise to ethical concerns. Finally, the data collection process can be difficult because the participant observer has to find a way to unobtrusively collect data from the other participants (Christoph 2004: 82).

15.4.5.4 Log Files

Computer-based systems commonly maintain log files that record events occurring during the operation of the system or an application for the purpose of diagnostics, surveillance, statistics, or improvement. In evaluation studies, log files can be used as an automatic data collection mechanism. While users are interacting with the system being evaluated, their behaviors, including commands, errors, messages, navigation paths, etc., are automatically captured and written to one or several log files or databases (Christoph 2004: 84f.).

Log files are an inexpensive and (similar to physiological measures, cf. Section 15.4.5.5) unobtrusive method for collecting data from users. However, they are not without problems (Christoph 2004: 85). First, the fact that data can be gathered unbeknownst to users raises ethical issues with respect to what subset of the user-related information that can be feasibly collected should in fact be collected. Second, logs can be used to accumulate data, but how should this data be analyzed to discover useful information, which may be buried deep within the data? To extract relationships and correlations hidden in the data, data mining techniques (Hand et al. 2001; Hearst 2003) can be useful. Third, low-level data in log files may support different interpretations of the user's behavior, so it may be necessary to (unobtrusively) collect further data for disambiguation from the user during the study, e.g. by adding auditory and visual input channels to the setup.

The information in log files can also be used directly to improve the performance of systems. For example, the exchanges between chatbots (cf. Chapter 11.3.2) and users can be logged to identify user inputs to which the bot did not have an (appropriate) response and to enhance the chatbot's database of response categories to enable the bot to handle these new cases.

15.4.5.5 Physiological Measures

Physiological data consists of variations in motor output and body readouts of human subjects that provide subconscious indicators of their emotional, mental, and physical state. Some kinds of physiological responses are readily observable by other people (e.g. facial expressions, voice inflection, gestures, etc.) while others, such as body temperature, perspiration, and blood pressure, can be detected through physical contact or measured by means of special instruments (cf. Chapter 13.1.1).

The attractiveness of collecting physiological data from participants comes from the fact that an individual has only limited conscious control over these responses. Hence, physiological measures are believed to generate more *objective* data than, for example, questionnaires (cf. Section 15.4.5.1) and observational studies (cf. Section 15.4.5.3), whose results may be affected by the self-selection and self-presentation of the participants and the subjectivity of the observer, respectively (Christoph 2004: 85).[256]

But evaluators should be careful when drawing conclusions about a particular construct on the basis of measurements obtained for the physiological variables that operationalize it (cf. Section 15.4.3). For example, participants' experience of stress indicated by physiological measures may not be due to the experimental intervention (e.g. being watched by an embodied interactive agent, cf. Chapter 10.4) but result from the experience of being subjected to physiological measurement, since the experimental apparatus necessary to collect physiological measures may create some kind of hospital or laboratory atmosphere that evokes negative associations and memories (Christoph 2004: 85f.).

15.4.6 Data Analysis

Data analysis (Christoph 2004: 92ff.; Hardy & Bryman 2004) processes the data collected in an evaluation study to extract information and draw conclusions from the data ([INT 201]). Ideally, all stages of the evaluation research (study design, data collection, data analysis, report writing and review) involve consideration of data analysis to some extent (Wisler et al. 1992). The results of data analysis are used to answer the research question (cf. Section 15.4.2). The answer should include conclusions from and explanations for the results, any generalizations that can be made from the results, influencing factors that might explain unexpected results, and questions that could be addressed by future research (Christoph 2004: 95f.).

---

[256] Prendinger et al. (2007) used an eye tracker to collect quantitative data about their subjects' focus of attention indicated by their eye movements during interactions with a life-like animated interface agent. They argued that "[t]he tracking of eye movements lends itself to reliably capturing the moment-to-moment experience of interface users, which is hard to assess by using post-experiment questionnaires" (p. 282). Prendinger et al. found that the agent's deictic gestures (cf. Chapter 10.1.2) more effectively guided participants' attention toward relevant parts of the interface than a text box or spoken comment.

Depending on the type of data being analyzed, two major types of data analysis can be distinguished: qualitative analysis is concerned with non-numerical (qualitative) data, whereas quantitative analysis processes numerical (quantitative data).

15.4.6.1 Qualitative Data Analysis

Qualitative data analysis (Miles & Huberman 1994; Denzin & Lincoln 2003; Hardy & Bryman 2004: Part V; Silverman 2005) examines and interprets any non-numerical data collected during an evaluation study, including the textual, narrative, and visual data generated by informal observations (cf. Section 15.4.5.3), unstructured interviews (cf. Section 15.4.5.2), focus group transcripts (cf. Section 15.4.5.2), document analysis, etc. (Morra-Imas & Rist 2002). The goal is to discover common words, phrases, issues, themes, or patterns (e.g. changes over time, causal relationships, etc.) in the data (Morra-Imas & Rist 2002), to identify connections between them, and to formulate conclusions and explanations with respect to the categories and relationships found in the data. Approaches to qualitative data analysis include *grounded theory* (Strauss 1987; Glaser 1992; Charmaz 2006), *semiotics* (Eco 1979; Chandler 2002; Deely 2005), and *conversation analysis* (Sacks 1995; Hutchby & Wooffitt 1998; ten Have 1999). Important tasks in qualitative data analysis are listed below (Miles & Huberman 1994: 44; Beusch & Zäch 1997):

- *Coding*. Assigning descriptive labels (*codes*) to segments of the data;
- *Search and retrieval*. Locating and retrieving coded segments from the data;
- *Aggregation*. Developing codes into superordinate codes or categories;
- *Data linking*. Establishing relationships between coded parts of the data;
- *Data display*. Representing data, codes, categories, memos, and their relationships;
- *Memoing*. Recording ideas, questions, notes, etc. about data, codes, and categories;
- *Theory building*. Developing concepts and theories based on the network of codes, categories, and memos.

Seidel (1998) presented a model of qualitative data analysis that is shown in Figure 93. According to Seidel, the analysis of qualitative data involves “noticing things,” “thinking about things,” and “collecting things” as activities in a non-linear process that is *iterative and progressive* (the analysis process is cyclic and moves upward like a spiral), *recursive* (a given activity can take the analysis back to a previous activity), and *holographic* (the entire analysis process is involved in each step).

The qualitative approach is appropriate for the early stages of research projects ([INT 202]) when little is known about the target domain and population since its open-ended nature makes it particularly suitable for exploratory research. In contrast to quantitative studies, which use tools such as questionnaires, sensory equipment, and statistics software to collect and analyze numerical data, in a qualitative study, the investigator is the primary instrument for gathering and processing data ([INT 202]) from interviews, observations, focus groups, etc. (although different computer-based tools to facilitate the tasks of qualitative data collection and analysis are available). The data that can be collected and analyzed in a qualitative study is much more varied and detailed than numbers and statistics ([INT 202]). The research process is also much less pre-structured (and hence more flexible but also less predictable) than in quantitative studies. On the downside, because qualitative research relies heavily on people for data collection and analysis and qualitative methods are commonly labor-intensive and time-consuming (Morra-Imas & Rist 2002), the number of participants in a study has to be kept

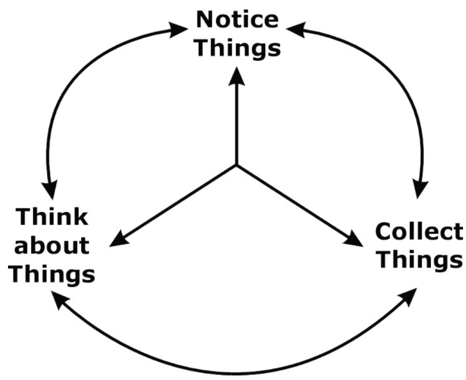

Figure 93. A model of qualitative data analysis. Adapted from Seidel (1998, Figure 1).

small. Furthermore, the inevitable subjectivity of investigators must be addressed to ensure the validity of the data and the analysis. Finally, due to the small sample sizes involved, it is usually more difficult to make generalizations from the results of qualitative data analysis ([INT 202]).

15.4.6.2 Quantitative Data Analysis

Quantitative data analysis (Wisler et al. 1992; Blaikie 2003) is concerned with the transformation of *numerical* data, which can be counted, classified, measured, and sorted using formal, objective methods. The data may come from a questionnaire (cf. Section 15.4.5.1), structured interviews (cf. Section 15.4.5.2), formal observations (cf. Section 15.4.5.3), analysis of log files (cf. Section 15.4.5.4), and/or physiological measures (cf. Section 15.4.5.5). Analysis of quantitative data is performed using two major kinds of *statistics*[257] (Coolidge 2006; Triola 2006): descriptive and inferential statistics (Morra-Imas & Rist 2002; Christoph 2004: 92ff.).

As a first step, *descriptive statistics* summarizes and presents the collected data in terms of graphs, tables, and calculated numerical values ([INT 203]). It shows the *distribution* of variables (i.e. the frequency of occurrence of their values) using frequency tables or histograms (Wisler et al. 1992). The distributions of the different variables in the data set are summarized by calculating measures of *central tendency*, which describe typical values of variables ([INT 204]), including the *arithmetic mean* (average), *median* (mid-point), and *mode* (most frequent value) (Morra-Imas & Rist 2002), and measures of *dispersion* (or *variability*), which indicate the scattering (or diversity) of the data ([INT 205]). One very popular way to measure the dispersion of interval or ratio data (cf. Section 15.4.5) is the *standard deviation (SD)*, whose value represents the average distance of a set of values from the mean ([INT 206]). The higher the standard deviation, the more deviation in the data, and vice versa (Christoph 2004: 93; [INT 207]). Measures of relationship (or *association*) indicate the strength of the relationship between variables (Morra-Imas & Rist 2002). *Correlation* is a measure of association that shows both the strength and the direction of the relationship between two variables (Chen & Krauss 2004: 1035). Two variables may change in the same direction (*direct relationship*) or in opposite directions (*inverse relationship*) (Morra-Imas & Rist 2002).

[257] Statistics is the branch of applied mathematics concerned with the collection, analysis, interpretation, and presentation of numerical data in large quantities and the use of probability theory to make generalizations or predictions from the data ([INT 209]).

Inferential statistics uses known statistical descriptions of data about a random sample of subjects to infer information about certain unknown aspects, including assumptions about the overall population (*generalization*) and *predictions* about future observations (Heffner 2004b; [INT 208]). Two important aspects of inferential statistics include parameter estimation and hypothesis testing (using statistical tests). *Parameter estimation* performs calculations on data collected from a representative sample consisting of randomly selected members of a statistical population (cf. Section 15.2.2) to estimate quantities (*parameters*) that describe this population. Any number that characterizes the distribution of a variable measured on a given population is a potential population parameter (Wisler et al. 1992). Examples include the mean and the standard deviation. Estimates for population parameters may be single 'best' values (*point estimates*), or they may consist of the upper and lower boundaries of an interval (*interval estimates*), which represent *confidence limits* defining an estimated range of values that contains the unknown value of a particular population parameter with a certain probability (the *confidence level*) (Wisler et al. 1992; [INT 210]).

Statistical *hypothesis testing* involves comparing the likelihoods of the *null hypothesis* (the observations have occurred by random chance) and its competitor, the *alternative hypothesis* (the observed effect is real) ([INT 211]; [INT 212]). If the probability p that the observed effect will be obtained if the null hypothesis is true (called the *p-value* or the *level of significance*) is small ($p < 0.05$), the null hypothesis is rejected because there is at least a 95% certainty that the observed effect did not occur by chance, and the effect is called *statistically significant*. Otherwise, it is *statistically non-significant* ([INT 212]). However, a significant effect is of no real practical value if the observed difference between two groups or conditions is small. In other words, the *effect size*, rather than the effect alone, is important (Christoph 2004: 93).

The major advantage of quantitative data analysis is that it generates reliable and valid ('objective') results in the form of numbers and statistics by means of well-defined and established mathematical methods (provided that the input data is also reliable and valid; otherwise, it is "garbage in – garbage out"). The calculations involved in the analysis are nowadays performed by computers, which increases efficiency and makes the analysis of large samples feasible. The same data set can be subjected to different kinds of statistical analysis ([INT 209]). The obvious disadvantage of quantitative analysis methods is that they are not applicable to data that does not have a numerical format, such as the data generated by informal observations (cf. Section 15.4.5.3), unstructured interviews (cf. Section 15.4.5.2), and focus groups (cf. Section 15.4.5.2). In general, quantitative approaches should be used in later stages of research projects, when the goal and object of the research are clear beforehand, precise measurements and analyses of constructs are aimed for, and all details of the study can be worked out before the data collection stage ([INT 202]).

15.5 Guidelines for Agent Evaluation

The evaluation of embodied interactive agents is a core activity of the development process. Given the many opportunities for making less than optimal design decisions, careful and systematic investigation of the precise nature and benefits of different (characteristics of) embodied interactive software agents in different contexts and with different users becomes an essential task in order to reduce the number of ad-hoc choices based on the intuitions and opinions of a small group of developers or even a single person. However, the evaluation of embodied interactive agents is also not an easy task. Various issues can be identified:

- As the research field of embodied interactive agents is still quite young, researchers have focused more on building agents and developing the technologies involved than on evaluating them. In publications on embodied interactive agents, evaluation has tended to be an afterthought, which was described only at the end of the paper (Ruttkay et al. 2004: 30). Large-scale evaluations have been quite rare (and still are).
- Isbister and Doyle pointed out that the field lacks formal theories of embodied interactive agents as well as standard criteria for evaluating them. The fundamental research problem has not yet been formulated in a way that is accepted by (almost) everybody, and the same is true of the definitions of central concepts, including 'believable,' 'embodied,' 'social,' and 'conversational.' The various contributing research communities (cf. Table 30) have different views of the nature of embodied interactive agents and what qualities play a role in their evaluation (Isbister & Doyle 2004: 5f.).
- Embodied agents have only recently begun to see commercial application on a larger scale, e.g. as chatbots on the World Wide Web (cf. Chapter 11.3.2). Consequently, feedback from end-users and the market on the strong points and deficiencies of agent-based applications has been sparse.
- The diversity of designs, tools, terminologies, settings, and methods makes it difficult to compare the outcomes of different evaluation studies. As a result, evaluations of embodied interactive agents often seem to produce inconsistent results with respect to the effectiveness of agents and their components while the actual problem is that the studies concerned differ in so many respects that they cannot be compared (Berry et al. 2005: 306).
- According to Sanders and Scholtz (2000: 3), evaluation is more difficult for interactive systems (including embodied agents) than for other kinds of software. Sanders and Scholtz provided the following reasons for this (p. 3). First, the nature and priority of performance metrics depends on the goals that the user wants to achieve with the system. Second, users differ and so do their needs and their expectations of the system. Third, from the user's point of view, the system being evaluated is more than the sum of its components. It is possible to design a badly received system made up of high-quality components but also to compensate for weak components with an effective interface design. Evaluation has to address both the system as a whole (cf. Section 15.3.4) and the effects of individual components (cf. Section 15.3.5). Finally, as interactive systems grow in complexity, so does their evaluation. Enhancing a system with speech and dialogue capabilities and embodying the interface might simplify the interaction with the user, but evaluation complexity increases.
- Embodied interactive agents are often designed to play their part in face-to-face dialogues with the user (cf. Chapter 10.1 and Chapter 11). However, embodied communication is very complex and still not completely understood, neither in the human-human nor in the human-agent context (Ruttkay et al. 2004: 40). Hence, both the design and the evaluation of embodied interactive software agents have to be based on incomplete theories of human communication (cf. Section 15.3.5). Comparing agents to humans according to a partial account of human communicative abilities is unlikely to produce definitive results that can inform the design of human-agent communication. The same is true of current models of emotion and personality (cf. Chapter 13), language processing (cf. Chapter 11.1), expertise (cf. Chapter 12), etc.
- The empirical evaluation of embodied interactive agents requires a methodologically sound approach. The problem with early evaluation studies was that they often did not meet the standards of empirical research methodology (Dehn & van Mulken 2000: 17ff.). Problems found in these studies include:

- In system-level evaluations (cf. Section 15.3.4), the effects of the agent were not clearly separated from those of the application in which the agent was embedded. Micro evaluations (cf. Section 15.3.5) that addressed the effects of certain aspects of an agent did not distinguish them from the effects of other aspects that were not investigated.
- Experimental and control conditions (cf. Section 15.4.4.2) differed in several aspects, not just the one to be studied, and in some cases, there was no control group.
- Studies used vague notions like 'believability' or 'trustworthiness' as constructs (cf. Section 15.4.3) and either did not define them at all or gave their own definitions,[258] which did not necessarily concur with those of other researchers (Ruttkay et al. 2004: 29f.).
- Variables were tested only on a small set of items (e.g. a single instance of a female and a single instance of a male character to assess how the gender of an embodied agents affects the user), which made it impossible to control for random variations in the items (e.g. hair style, facial expression, etc.) that might influence user ratings.
- The studies failed to demonstrate the validity of the variables defined to measure the constructs adopted for the study.
- Participants in evaluation studies were typically recruited from among university or college students and/or staff because researchers had convenient access to them. However, those individuals were not necessarily representative of the target group of users for the agent being evaluated.

Given these issues and based on the discussion in the previous sections of this chapter, a number of guidelines for evaluating embodied interactive software agents can be formulated:

- *Iterative evaluation and design*. Design and evaluate your embodied interactive agent iteratively, performing formative evaluations early and regularly throughout the development process while not neglecting the effectiveness of the overall user-agent system (Johnson 2004). Take advantage of the complementary contributions from expert reviews and user studies in your formative evaluations. Use the results of consecutive evaluations of designs, prototypes, and alternative versions of the agent to successively improve the products of the development process.
- *Focus on user-agent interaction*. Evaluate users' interaction with the agent rather than only the agent (Johnson 2004). Have subjects *interact* with the agent and not merely watch a presentation given by the agent because the true potential of embodied interactive agents lies in their interactional capabilities, in particular their capacity for emulating face-to-face communication (Dehn & van Mulken 2000: 19, cf. Chapter 7.1). Use these evaluations to identify both factors that hinder smooth user-agent interaction and factors that promote successful interaction, and eliminate or exploit them, respectively (Johnson 2004).
- *Separate the agent from the application*. Make sure that the agent and not the application is evaluated by separating the effects of the agent from those of the application (Ruttkay et al.

[258] One attempt to provide succinct definitions of constructs like believability, efficiency, helpfulness, learnability, and trust for the evaluation of embodied interactive agents was made by Ruttkay et al. (2002). However, quite a number of these definitions remain vague. For example, Ruttkay et al. stated that an agent is believable "if it acts according to the expectations of the user." Apart from the fact that this definition does not mention the "illusion of life" criterion (Bates 1994) found in many other definitions of believability, it is not clear what the "expectations of the user" are, although they are the central component of the definition.

2004: 43f.). (However, keep in mind that in the eyes of the users, the effectiveness of an agent also depends on the effectiveness of the system in which it is embedded.) If the evaluation targets particular aspects of the agent, it is necessary to separate the effects of those aspects from effects of other aspects that are not under investigation.

- *Research question*. Formulate a clear and focused research question. Broad questions like whether the use of embodied interactive software agents is beneficial or not cannot be answered, neither with a single evaluation study nor in general, due to the diversity of the factors involved (cf. Section 15.2). Clearly state the goal of your research (understanding the effects of embodied interactive agents, comparing alternative agent designs, assessing the performance of component technologies, determining conformity to standards, etc.) and who will benefit from it (cf. Section 15.1), and how.
- *Constructs*. Provide clear descriptions of the constructs and their operationalization. Avoid vague constructs like 'believability,' 'effectiveness,' or 'trustworthiness.' Prefer established measures to allow others to understand and replicate your results (Isbister & Doyle 2004: 9). Demonstrate the validity of the variables used to measure effects (Dehn & van Mulken 2000: 18). In general, there is a need for standardized evaluation measurements for embodied interactive agents (Isbister & Doyle 2004: 6).
- *Selection of subjects*. Choose the participants for the study carefully. Those who are conveniently available may not be the best choice if they are not representative of the agent's target group of users. Be sure to match the experimental and control groups on all important aspects (except for the one being investigated, of course). Developers should not be involved as subjects in testing and evaluation because they are naturally biased when it comes to judging the qualities of their brainchild (Isbister 2006: 267). Still, they should watch tests that involve users from the target group interacting with their agent, either live or later on video, because "[s]eeing is believing" (p. 269).
- *Novelty effect*. Beware of the novelty effect and how it influences users' judgment and performance (Ruttkay et al. 2004: 50). Given that embodied interactive agents are not as common as other types of software systems yet, it is quite likely that they will be a new technology for the participants in the study. When users are exposed to an embodied interactive agent for the first time, their increased interest in the new technology may initially cause them to perform better and to rate the agent more positively ([INT 213]), but these effects are likely to fade as the agent ceases to be new and surprising and becomes familiar and predictable. Hence, reliable information about the *long-term effects* of agents cannot be obtained by means of studies that have users interact with an agent only once and for a short time.
- *Random assignment to varied conditions*. When conducting experimental research, assign users randomly to a range of different conditions (cf. Section 15.4.4.2). The relative and comparative nature of hypotheses on embodied interactive agents forbids evaluating a given agent on its own. Instead, the study should compare the agent to one or several different versions of the agent or to other agents, an interface without an embodied interactive agent, or a human being. What condition(s) the user is exposed to (and in what sequence) should be decided by random chance rather than by the user or the experimenter (Nass et al. 2000a: 375ff.).
- *Control group*. In experimental studies, be sure to include a baseline for comparison in the form of an appropriate control group that does not receive the manipulation of the experimental condition. Design the experimental condition and the control condition in such a way that they only differ in terms of the dimension under investigation to make sure that any effects found are attributable to differences on that dimension rather than other factors.

For example, compare an agent with a 2D cartoon character representation to a 3D cartoon agent rather than a 2D cartoon to a 3D photorealistic character (Ruttkay et al. 2004: 62, cf. Chapter 9.5). Based on the results of such controlled comparisons, specific guidelines for the design of embodied interactive agents can be formulated (cf. Section 15.1 and Section 15.4.4.2).

- *Economy*. Provide for inexpensive evaluation (Sanders & Scholtz 2000: 14). Limit the participants in the study to a feasible number. More participants can mean more accurate results, but they also mean more work during data collection and analysis. Keep the number of constructs and the variables for measuring them to the necessary minimum. Prefer metrics that are easy to measure and do not require costly manual analysis (p. 14). Collect as much data as possible automatically using, for example, log files or sensory equipment (cf. Section 15.4.5.4 and Section 15.4.5.5) (p. 14). Use quantitative rather than the more labor-intensive qualitative data analysis techniques where possible (cf. Section 15.4.6). Evaluations that are inexpensive to do will be performed more often, with larger numbers of users (which may be distributed across several evaluation runs), and with several different versions of an agent or its components (p. 14).
- *Ethical dimension*. Embodied interactive agents have the potential to form social-emotional relationships with users (cf. Chapter 7.16) and to influence their attitudes and behaviors (Fogg 2003). It follows that agent developers are morally (and possibly someday also legally) responsible for their 'offspring' and its actions. For example, a chatbot (cf. Chapter 11.3.2) for children that uses abusive, racist, sexist, etc. language or tries to attract its young interlocutors to drugs or to buying the products of its employer gives rise to serious ethical concerns.[259] Therefore, embodied interactive software agents should also be evaluated with respect to their ethical code, not just their technical or task-related capabilities (Barker 2006).
- *Use of evaluation results*. Know how and when to apply the results of evaluation studies and who to entrust with their implementation. Shape the culture within the development team to ensure the openness of team members to constructive criticism based on evaluation research (Isbister 2006: 268f.).
- *Postproduction evaluation*. Do not deploy an agent and then forget about it. Instead, collect information about its effects in the real world to inform the design of future versions of the agent or the design of other agents. Examples include user and peer reviews, reported annoyances, "fan mail," etc. (Isbister 2006: 274).

15.6 Summary

The design of embodied interactive software agent depends on evaluation, i.e. the goal-directed, methodical, and systematic process of collecting, analyzing, and interpreting data about the effects of a particular agent in a particular state of development on a particular group of users in a particular context. Drawing on the extensive existing knowledge of evaluation methodology,

---

[259] Interestingly, an incident was reported on the web site of a German computer magazine in December 2007, where a Santa Claus chatbot employed by the world's largest software company had to be 'retired' because it involved minors in a conversation about certain sexual practices and its developers were unable to remove the parts responsible for this inappropriate behavior from its programming ([INT 214]).

this chapter reviewed various methods, strategies, and techniques for conducting different kinds of evaluation research and, based on this review, compiled a list of guidelines for the evaluation of embodied interactive agents.

Motivations for doing evaluation research on and with embodied interactive agents were identified in Section 15.1. In general, evaluation helps developers to focus their efforts, identify progress, make the best choice, and determine success (Sanders & Scholtz 2000: 1). In addition, evaluation research can produce guidelines for the design of embodied interactive agents and their components. In the next section, it was argued that the evaluation of these agents is influenced by three categories of factors: design parameters of the agent, characteristics of the user, and aspects of the context of use. Each of these categories was discussed in detail in Section 15.2.

Section 15.3 described different types of evaluation research that differ with respect to the time and level at which the evaluation is carried out. First, Section 15.3.1 outlined the stages of the software development process and presented several software life cycle models. The next two sections covered process-oriented, exploratory (formative) evaluation (cf. Section 15.3.2) and outcome-oriented (summative) evaluation (cf. Section 15.3.3). Macro (or system-level) evaluation (cf. Section 15.3.4) is concerned with the overall system consisting of the agent, the user, the application, and the context of use. In contrast, micro evaluation (cf. Section 15.3.5) investigates whether an agent with certain attributes meets expectations, and further identifies configurations of design parameters that achieve desired attributes of agents.

Section 15.4 reviewed evaluation research methodology. Fundamental concepts include the research model (an overview of the main steps involved in the research, cf. Section 15.4.1), the research question (the question that the research purports to answer, cf. Section 15.4.2), and the (psychological) construct (a latent mental entity that cannot be measured directly and therefore has to be operationalized in terms of variables, cf. Section 15.4.3). The research strategy (cf. Section 15.4.4) provides a plan for the conducting the research. A broad distinction was made between empirical research, which collects data first-hand from experience, observation, or experimentation, and desk research, which reviews available material from the perspective of a different research question. Three empirical research strategies were reviewed: surveys (cf. Section 15.4.4.1), experiments (cf. Section 15.4.4.2), and case studies (cf. Section 15.4.4.3). Desk research was discussed in Section 15.4.4.4. For each of these research strategies, different methods can be used to collect data (cf. Table 31). The various data collection methods commonly used in empirical research were reviewed in Section 15.4.5. They include questionnaires (cf. Section 15.4.5.1), interviews (cf. Section 15.4.5.2), observations (cf. Section 15.4.5.3), log files (cf. Section 15.4.5.4), and physiological measures (cf. Section 15.4.5.5). The process of data collection generates descriptive, non-numerical and/or numerical data, from which the data analysis stage (cf. Section 15.4.6) extracts information, draws conclusions, and makes generalizations. Qualitative data analysis (cf. Section 15.4.6.1) examines and interprets non-numerical (textual, narrative, and visual) data, which involves collecting, noticing, and thinking about items in a process that is iterative and progressive, recursive, and holographic. In contrast, quantitative data analysis (cf. Section 15.4.6.2) deals with numerical data, which it processes using the methods of descriptive and inferential statistics.

Finally, Section 15.5 derived a set of guidelines for designing and conducting evaluation studies of embodied interactive software agents. These guidelines are summarized below:

- Design and evaluate the embodied interactive agent iteratively.
- Focus on user-agent interaction.
- Separate the agent from the application.

- Formulate a clear and focused research question.
- Provide clear descriptions of the constructs and their operationalization.
- Choose the participants for the study carefully.
- Beware of the novelty effect and how it influences users' judgment and performance.
- In experiments, assign users randomly to varied conditions.
- In experiments, include a control group to have a baseline for comparison.
- Prefer inexpensive ways to perform evaluations.
- Evaluate agents with respect to their ethical code, not just their capabilities.
- Have a plan for the use of evaluation results.
- Keep evaluating after the agent has been deployed.

16 Pedagogical Agent Roles on the VLC

The high priority placed by most educational theorists on the social construction of knowledge has provided a powerful new vehicle for the expansion of research and development in the area of software agents. Unless the socially interactive aspects of face-to-face learning environments can be shown to have a digital equivalent, online courses, while convenient for many students, will inevitably be regarded by a large number of educators as second best. (Dowling 2000: 44).

The previous chapters discussed the design of different components of embodied pedagogical software agents in detail, including embodiment, behavior, conversation, expertise, and emotion and personality. Two further chapters were concerned with agent architectures and technical platforms and with the evaluation of embodied interactive software agents, respectively. In this penultimate chapter of the thesis, the insights from the previous chapters are applied to a discussion of different pedagogical agent roles in a particular scenario provided by the linguistic fieldwork classes on the Virtual Linguistics Campus. First, Section 16.1 outlines the current setup of the linguistic fieldwork classes. Section 16.2 then identifies and describes three potential roles for pedagogical agents to assist the learner in this scenario: the virtual native speaker, the coach, and the peer. Section 16.3 discusses the design of one of these roles (the virtual native speaker) in detail. Finally, Section 16.4 summarizes the chapter.

16.1 The Linguistic Fieldwork Classes

Linguistic fieldwork (Crowley 2007) involves the in-situ collection of raw linguistic data from speakers of unknown (and typically unexplored) languages, followed by the analysis of the collected data in order to uncover the linguistic structures (e.g. sound inventory, word and sentence structure, and vocabulary) of these languages. In general, fieldwork settings can provide authentic, explorative learning experiences for students of linguistics. Learners actively construct knowledge (cf. Chapter 2.2.3) about how to do linguistic fieldwork and about the target language as they uncover more and more of its *grammar* (i.e. the native speaker's subconscious knowledge about his or her first language). Unfortunately, for many learners and institutions, going to some linguistically unexplored area to do linguistic fieldwork is neither feasible nor affordable.

The linguistic fieldwork classes on the Virtual Linguistics Campus have been created to bring linguistic fieldwork scenarios to learners, using the interactive and multimedia features of the VLC (Unger 2006b). In October 2007, fieldwork classes were available for the following languages: Arabic, Bulgarian, Chinese, Georgian, German, Hindi, Hungarian, Japanese, Polish, Romanian, and Welsh (Cymraeg). Classes for Italian, Korean, and Turkish are currently under construction.

As their name implies, the linguistic fieldwork classes are *not* language courses. Students will not learn the language they are investigating, but they will learn *about* the language as they explore its structure using the methods of linguistic fieldwork. The emphasis of the fieldwork classes is on providing an environment in which learners can apply and develop their linguistic analysis skills in a fieldwork setting that is as authentic as it can be created with the capabilities of a state-of-the-art e-learning platform. Learners' construction of knowledge about the target language is a useful side effect but not the primary goal of a fieldwork class on the VLC.

Each linguistic fieldwork class consists of the same basic set of units in the same sequence (cf. Table 32), which are made available (activated) successively (Handke 2006a: 25f.). Of

Table 32. The units of a linguistic fieldwork class on the VLC. Compiled from Unger (2006b: 172–175).

Unit Title	Central Topics
Class Preliminaries	General information about how to study the fieldwork class online
Fieldwork Methodology	Background on methods and tools for linguistic analysis; introduction of the (real) native speaker; general information about the target language
Simple Objects	The sound system of the language (i.e. its vowels and consonants)
Number	More on the sound system; formation of the plural and the dual of nouns with and without a determiner
Adjectives	Use of adjectives in attributive and in predicative function
Adpositions	Structure of adpositional phrases (i.e. noun phrase plus preposition or postposition)
Simple Sentences	Structure of sentences with intransitive and monotransitive verbs (i.e. verbs without and with a single object)
Time and Tense	Representation of time relationships (past, present, and future, both simple and progressive)
Interrogatives	Structure of simple interrogative sentences (yes-no and wh-questions)
More on Sentences	Structure of more complex sentences (e.g. sentences with ditransitive verbs (two objects), semantic-role relationships, etc.)
Summary	Summarization of all observations made in the class; identification of gaps, loose ends, and areas for further study; vocabulary test

course, the contents of the individual units are specific to the language that is the topic of the fieldwork class. Each unit builds on its predecessor (if available) and begins with a cumulative summary of the hypotheses that can be formed about the target language on the basis of the data collected up to this point. Learners compare these hypotheses to their own, which they built during their previous explorations and wrote down in the obligatory Workbook of the class (Unger 2006a).

The core of a typical fieldwork unit is the *virtual room* (shown in Figure 94), which provides the setting for learners to listen to and explore the language. The central element of this room is the *virtual native speaker*, an animated character playing the role of the learner's informant (i.e. primary source of linguistic data) about the language being explored. Using a virtual instead of a real native speaker in the linguistic fieldwork classes makes it possible to address a number of problems with human informants:

- They are not always available when the learner has time to do fieldwork.
- They may be less than fully cooperative and might even try to deceive the learner.
- They may lose their patience when the learner asks them to repeat things (again and again).
- They are sensitive to, and may be unforgiving of, faux pas in the learner's behavior (even inadvertent ones).
- They are unpredictable in their linguistic and other behaviors, so eliciting the required data about specific linguistic phenomena from them may take considerable time and skills, which (beginning) learners of linguistic fieldwork may not have.

In contrast, the virtual native speaker is always available, cooperative, patient, tolerant, and predictable. When the learner enters the virtual room, he or she finds the speaker sitting behind a table and ready for the learner to start exploring. The virtual native speaker is surrounded by a number of objects in the room. When the learner clicks on one of these objects, the speaker

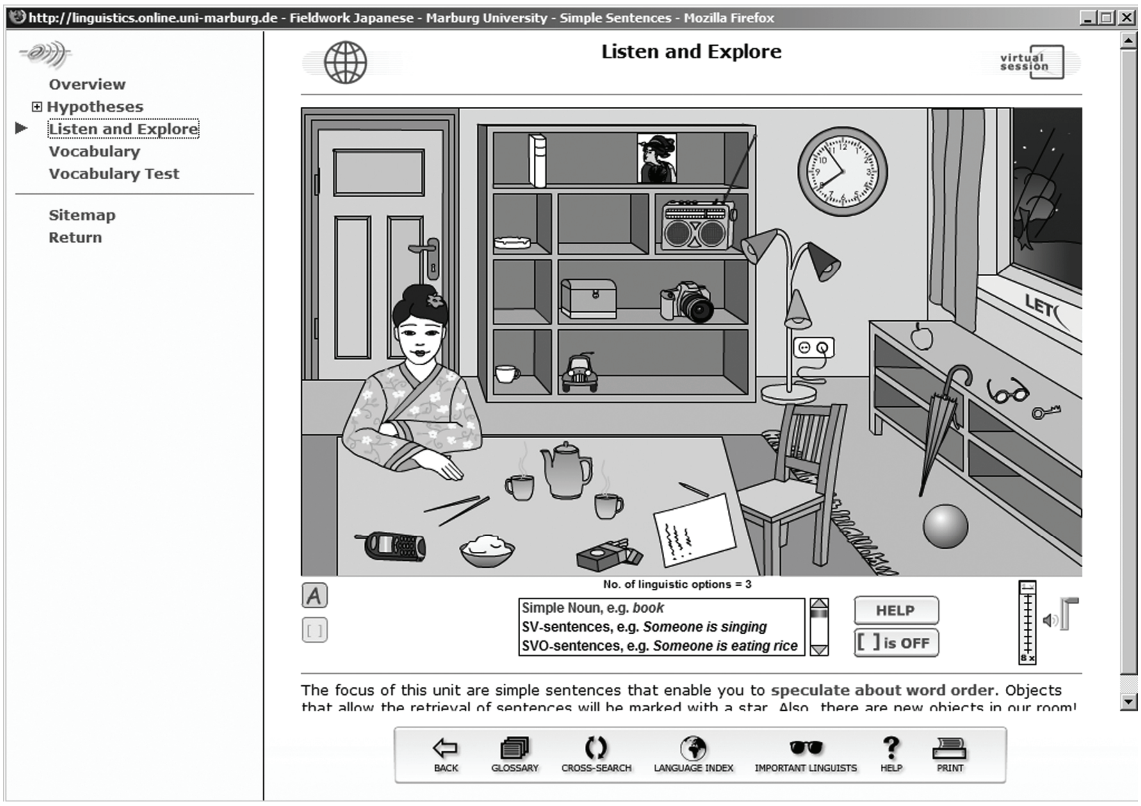

Figure 94. The virtual room of the fieldwork class Japanese on the VLC, featuring the virtual native speaker Reiko and various clickable objects. Screenshot taken on 6 December 2007.

produces the word for the object in its[260] native language or uses the word in a more complex phrase or sentence, depending on what type of linguistic structure was selected for investigation by the learner from a list of options. Each utterance of the virtual native speaker is accompanied by an English gloss, its representation in the writing system of the language, and its phonetic transcription (if enabled). In later units dealing with more complex actions and relationships in the language, a second virtual actor is added to the room where required, which is from the same culture but of a different sex than the speaker. This actor does not speak but only acts out the virtual native speaker's utterance.

Learners collect linguistic data from the virtual native speaker and from the environment and use this data to speculate about various structural aspects of the language, including the sound system, patterns of plural formation, the use of adjectives, whether the language has prepositions or postpositions, the structure of simple and more complex declarative and interrogative sentences, the representation of time relationships, etc. Each unit of a linguistic fieldwork class focuses on one of these aspects, starting with the sound system and successively moving toward higher structures, but the learner's knowledge of previously encountered aspects may be consolidated and further developed as well, depending on what additional linguistic data is presented in the unit. Along the way, learners build up a small vocabulary of words in the target language.

[260] The neuter form of reference to the virtual native speaker is deliberately used here and throughout the remainder of the chapter to emphasize that 'it' is not a real speaker but an artificial character.

There is no script for learners' exploration of the virtual room or their interaction with the virtual native speaker, apart from some general instructions and a broad description of the task to be accomplished by the learner, such as "Find out how Japanese constructs interrogative structures." Learners decide for themselves how they explore the virtual room to complete the task, and how often; there are no pre-defined learning paths through a fieldwork unit (although the overall sequence of the units in a fieldwork class is fixed). Through the coordinated use of auditory and visual elements in the virtual room, learners can easily establish the connection between an object, relationship, or action on the one hand and the virtual native speaker's utterance on the other. The material used consists of authentic and concrete linguistic samples recorded from a real native speaker of the language. A help screen for the current virtual room is available on demand. Further information about linguistic topics can be accessed through the options in the menu bar (Cross-Searcher, Glossary, and Language Index, cf. Chapter 2.4.1).

For languages with an alphabetic writing system (which are easier to handle given a Western keyboard layout), some fieldwork units integrate a vocabulary test that allows learners to check how many words from the previous units they already know. The test takes place in a different virtual room that is modeled after a classroom and features two virtual actors. One of them is the virtual native speaker of the fieldwork class, the other a native speaker of English. The procedure of the vocabulary test is as follows: the virtual native speaker produces a word in its native language. Additionally, the unknown word is presented in its written and transcribed form on a card behind the virtual native speaker. Furthermore, the card provides the option to listen to the word again. Next, the learner has to type the English translation of this word into an input box. The solution is pronounced by the speaker of English and additionally presented both in spelling and transcription on a card behind that speaker, also with the option to listen to the word again. Depending on the setting, words that were not correct will be presented again in the current round.

The linguistic fieldwork classes have proven a valuable addition to the Virtual Linguistics Campus, not only because they enrich the VLC's repertoire of courses but also because they take a constructivist approach (cf. Chapter 2.2.3) which differs from other courses on the VLC that focus more on the delivery of content. Each fieldwork class provides authentic linguistic material and offers learners the freedom and the challenge to actually *do* linguistics (rather than just read about it). During this process, learners formulate hypotheses which may or may not turn out to be correct. Incorrect hypotheses, revealed by the summary at the beginning of the following unit, encourage the learner to reflect on his or her analyses and knowledge and make adjustments where necessary.

16.2 Agents for Linguistic Fieldwork

While the linguistic fieldwork classes on the VLC are already a promising concept, there is room for improvement, in particular with respect to the social dimension of learning and the support provided by the environment for the individual learner. A general problem with the current format of the fieldwork classes is that learners by and large have to tackle the fieldwork tasks on their own. Although learners can get in contact with their instructor and peers through e-mail, the discussion forum, and the chat room (Unger 2006b: 170), their access to timely and individualized guidance, encouragement, companionship, and other kinds of support while they are doing field research in the virtual room is limited. Some of the shortcomings are analyzed in the following, and potential roles for embodied pedagogical software agents are identified.

16.2.1 Roles for Pedagogical Agents

Learners already have a companion in the virtual room when they study the units of a linguistic fieldwork class: the virtual native speaker. However, that character currently does not invite rich and flexible exchanges of the kind possible with human native speakers in real-world field situations because it can only react with pre-recorded utterances to the learner's point-and-click actions in the virtual room. The virtual native speaker lacks the linguistic competence to describe a wider range of situations in the room with its utterances, in particular unanticipated ones created by the learner or by another virtual actor, and to participate in interactions and performances with other virtual actors and with the learner. Furthermore, the speaker's options for bodily interactions with its environment are limited to a small set of idle behaviors (cf. Chapter 10.2), which are performed randomly to enhance the speaker's believability. If the virtual native speaker had a complete articulate body (cf. Chapter 9), it could not only use it to act on other virtual actors and the virtual room but also to complement its verbal utterances with non-verbal communicative behaviors involving combinations of facial expressions, gaze, gestures, posture, and locomotion (cf. Chapter 10.1), similar to a human speaker.

Another shortcoming of the virtual native speaker concerns its underexploited potential for building a relationship with the learner. The virtual native speaker is the one individual that learners (have to) interact with regularly throughout the entire fieldwork class. However, so far, the speaker does not even acknowledge the learner's presence in the virtual room, let alone remember the learner on subsequent visits or show any interest in him or her. Furthermore, since the virtual native speaker makes no effort to portray an interesting personality or to share details about itself, the learner has little opportunity to get to know the speaker as a person. In contrast, an enhanced virtual native speaker presenting itself as a 'deep' character and showing a perceptible interest in the learner may be more inviting as a regular interaction partner in the fieldwork experience.

Linguistic competence, non-verbal behavior, and relationship building are three areas in which the virtual native speaker can be improved to become a more competent, flexible, and interesting informant for learners. The role of the enhanced virtual native speaker is described in Section 16.2.2 below while detailed ideas for its design are presented in Section 16.3.

From the perspective of the cognitive apprenticeship model (cf. Chapter 2.2.3), the learners participating in a fieldwork class on the VLC can be regarded as apprentices in linguistic field research, who have different levels of proficiency and hence require different kinds and degrees of support. However, this apprenticeship is not complete without a master teacher coaching the learner while he or she is solving linguistic problems in the environment. So far, the fieldwork classes do not offer a personal coach for the individual learner. The virtual room gives learners freedom of exploration, but it does not provide guidance when the learner is unsure about how to proceed, encouragement when he or she is frustrated, or that one useful hint which helps the learner to get back on track after running into a dead-end with his or her analysis. Especially for learners who are new to linguistic fieldwork, the scaffolding and motivational support (cf. Chapter 12.3.2 and Chapter 12.3.8) given by a personal fieldwork coach could be useful. Since the (constant) availability of qualified human coaches for each learner is an issue, the second job opening for pedagogical agents in the fieldwork scenario is the role of coach or facilitator of learners' problem solving in the fieldwork process. This role is described in Section 16.2.3.

The third role of pedagogical agents in the linguistic fieldwork classes to be discussed in this chapter addresses the problem of the limited opportunities for social learning provided by the current virtual fieldwork environment (Unger 2006b: 170). Section 16.2.4 outlines a peer agent that explores the virtual room together with the learner. The peer role is inspired by the social-constructivist view of learning as a social and cooperative process in which learners construct knowledge through their interaction with others in their environment (cf. Chapter 2.2.3). Peer agents are designed to be similar to the learner in terms of age, status, and/or ability. Learners can cooperate or compete with them, enjoy their company, follow their example, avoid their mistakes, and teach them or be taught by them (cf. Chapter 12.3.10). These aspects of peer agents are addressed in detail in Section 16.2.4.

16.2.2 The Virtual Native Speaker

The virtual native speaker is a central character of the linguistic fieldwork classes because it provides learners with the linguistic input that allows them to speculate about the structure of the unknown target language. The speaker is a fixture in the virtual room in every fieldwork unit and the only individual that is always with the learner when he or she is exploring the room. In the course of a fieldwork class, the virtual native speaker and the learner have regular, extended sessions together over several weeks. Hence, there is ample opportunity for the speaker agent to engage the learner in interactions that expose him or her to more and more varied data from the target language and to build a social-emotional relationship with the learner (cf. Chapter 7.16), which preferably develops into a cooperative partnership. While the language barrier between the learner and the virtual native speaker is a problem for relationship building, there are still ways for the speaker to establish rapport with the learner. Relationship-building behaviors for the virtual native speaker are discussed in Section 16.3.2.4.

A positive relationship between the learner and the virtual native speaker is the prerequisite for a successful cooperation between the two in the virtual room. The virtual native speaker has a considerable responsibility in this cooperation because the learner relies on the speaker as the provider of linguistic data from the language under investigation. Therefore, the speaker agent has to be equipped with authentic knowledge of its native language. A minimum level of competence is currently available in the form of pre-recorded utterances that are linked to clickable objects in the virtual room and delivered by the virtual native speaker. While simple to implement, this 'stimulus-response' approach offers little flexibility in the interaction between the learner and the speaker. In contrast, the enhanced virtual native speaker is a conversational agent (cf. Chapter 11.3) which, similar to a human native speaker, has the ability to both produce and understand its native language at the level of complexity that is required for the domain of the virtual room, and can hold up its end of restricted domain-specific dialogues with other virtual actors and with the learner using that language. For example, a presentation team of two or more virtual native speakers may act out a particular scene in the virtual room, which can involve a short dialogue between the actors that illustrates particular phenomena of the target language (cf. Chapter 12.3.12). The performance and the dialogue may be scripted in advance or may be dynamically made up by the participants (cf. Chapter 14.2.4). Learners can watch these virtual performances and, importantly, become involved themselves, by asking questions about the performance (e.g. "What is X doing?," "Where is Y?," "How many Ys does Z have?") that elicit responses from the virtual actors involved; by making changes to the scene and collecting linguistic data from the reenacted performance; or by playing their own part in a performance and thus practicing and experimenting with structures of the language (cf. Chapter 12.3.12).

In addition, the speaker agent possesses knowledge about the entities in the virtual room and can use this knowledge to dynamically generate utterances expressing actions or relationships involving these entities. Ideally, learners can inquire about any entity (object, actor, or event) in the room and combine it with any other entity or entities in any way that makes sense in the language and in the environment of the virtual room, and the virtual native speaker produces the appropriate utterances. Furthermore, entities can be added to or removed from the room, and their position, activity, or status may change in the course of a session and even persist across sessions. All these manipulations can be performed not only by the learner but also by the virtual native speaker and other virtual actors. Thus, the virtual room becomes a dynamic environment, in which learners have options for experimentation and exploration that go beyond the small, pre-defined set of interactions that are currently possible in the room.

However, the space provided by the virtual room is still small and hence limits the learner's possibilities for exploration and interaction. As a potential future development, computer game technology can be used to extend the virtual room to a larger, three-dimensional simulated world for doing linguistic fieldwork, which contains different settings ('maps') and is populated by a number of embodied pedagogical agents playing the roles of virtual native speakers, fieldwork coaches (cf. Section 16.2.3), and peers (cf. Section 16.2.4), as well as by other human learners. Learners explore this virtual fieldwork world and interact with different human and artificial agents to accomplish a variety of linguistic fieldwork 'missions,' alone or in a team with other learners and peer agents.[261]

One significant obstacle to the interaction between the learner and the virtual native speaker is the language barrier. The learner knows very little about the speaker's native language at the beginning (in fact, uncovering the structure of that language is the point of a fieldwork class), and even later in the class, his or her knowledge of the language does not grow to a level that would allow him or her to participate in unrestricted everyday conversations. In turn, the virtual native speaker generally does not speak the learner's native language. To facilitate the interaction between the learner and the virtual native speaker, it is necessary to overcome or bypass the language barrier in some way. One possibility involves the use of non-verbal means of communication, including gestures, facial expressions, gaze, posture, and locomotion (cf. Chapter 10.1). In fact, the pointing gesture (using the mouse) is currently the only way for the learner to interact with the virtual native speaker. Further mouse-based gestures can be added, which allow learners to flexibly move, combine, or change entities in the virtual room, in order to increase the range of situations that the speaker agent can describe with its utterances. While current web-based learning environments offer learners few possibilities for non-verbal interaction apart from using the mouse, the virtual native speaker is portrayed as an animated embodied character and can use its face, hands, and body to perform various non-verbal communicative actions that complement its verbal utterances and otherwise support its interactions with the learner or enhance its believability as an individual. For example, to unambiguously refer to objects that are further away from its current location, the virtual native speaker could move through the virtual room toward the object of interest, and point at it or pick it up while producing its utterance.

---

[261] The model for this world is provided by the *Tactical Language and Culture Training System (TLCTS)* (Johnson et al. 2004b; Johnson et al. 2005a; Johnson et al. 2005b; Johnson 2007; Johnson & Valente 2008; Johnson et al. 2008), a simulated environment designed to help learners to develop the skills for accomplishing authentic communicative tasks in a foreign language and culture. Coached by a pedagogical agent, the *virtual aide*, the learner practices his or her communication skills by speaking the target language and choosing appropriate communicative gestures (cf. Chapter 10.1.2) in simulated social interactions with autonomous, animated native speakers of the language. The focus of these interactions is on spoken communication, (in-) appropriate non-verbal gestures, and cultural norms of etiquette and politeness (Johnson et al. 2004b).

A language is closely intertwined with the culture in which it is spoken. The virtual native speaker is a representative of that culture, which is reflected by its appearance, behavior, and interactions with others. Hence, apart from the language, learners can also construct useful knowledge about the culture from their interactions with the virtual native speaker and from observing (and participating in) the exchanges between several virtual actors. For example, interactions between actors of different status could reveal how the (power) relationships between the people portrayed manifest themselves in their utterances, which is particularly noticeable in languages like Japanese. Furthermore, the virtual native speaker's utterances can be accompanied by gestures, facial expressions, and other non-verbal behaviors, which learners can watch (repeatedly), relate to the verbal utterances, and compare to the use of non-verbal communication in their own language and culture. As a result, the learner's fieldwork addresses both the linguistic and the paralinguistic aspects of the target language, by analyzing and drawing conclusions from the integrated (verbal and non-verbal) performance of the virtual native speaker.

The previous paragraphs have portrayed the virtual native speaker as the learner's informant about the language he or she is exploring. But it is also possible to reverse the roles: the learner is a speaker of a native language (or several in the case of multilingual speakers) and can serve as the speaker agent's source of data from that unknown language. The result is a "tandem situation," in which both parties learn about the native language of the other. For the learner, this may be beneficial because it gives him or her first-hand experience of what it is like to be an informant. Furthermore, he or she can observe the virtual native speaker in its secondary role as linguistic field researcher and obtain insights into the structure of his or her own native language, which is subconscious rather than explicit knowledge for many native speakers. In addition, the virtual native speaker can also act as a role model for the learner that shows him or her how linguistic fieldwork should be done (cf. Section 16.2.4).

16.2.3 The Coach

Mace Windu: There's no doubt that the mysterious warrior was a Sith.
Master Yoda: Always two there are, no more, no less: a master and an apprentice.
Mace Windu: But which was destroyed? The master or the apprentice? ([INT 68]).

The coach implements the role of a master teacher in a cognitive apprenticeship relationship with the learner (cf. Chapter 2.2.3). A pedagogical agent in this role models various analysis and hypothesis formation skills involved in doing linguistic fieldwork for its apprentice (the learner) and then monitors and guides the learner as he or she is practicing these skills in the current fieldwork unit while working on the data supplied by the virtual native speaker, intervening at appropriate points to help the learner to solve the current problem (cf. Chapter 12.3).

The coach agent is embedded in a *linguistic workbench* environment, a potential extension of the current fieldwork-class setup that provides learners with tools and resources for collecting and organizing linguistic data from the virtual native speaker, performing different kinds of linguistic analysis on these data (phonological, morphological, syntactic, etc.), and formulating hypotheses about the structure of the target language (e.g. sound inventory, plural formation, word-order patterns, etc.) based on these results. Features of the linguistic workbench include:

- A linguistic keyboard for the phonetic transcription of utterances;
- Interactive chart diagrams for categorizing identified consonants and vowels;

- Vocabulary lists for organizing words and their translations;
- Mechanisms for tagging utterances with morphological, syntactic, and semantic information;
- Tools for decomposing words and sentences into their constituents and visualizing the resulting structures;
- Facilities for formulating, testing, and revising hypotheses;
- Record keeping of observations, analysis results, and hypotheses.

The assistance provided by the coach is intended to further the process of learners' linguistic problem solving in this environment. The coach typically does not deliver the final result of an analysis or the 'correct' conclusion from its results (this would undermine the constructivist approach of the fieldwork classes), but it provides the required scaffolding (cf. Chapter 12.3.2) for learners to accomplish tasks in linguistic fieldwork that he or she cannot (yet) do by himself or herself, always encouraging the learner to take the next step beyond his or her current abilities and ensuring that the learner manages to complete a given task successfully. The coach may supplant (take over for) the learner in steps which are necessary to complete the task but require skills that are too complex for the learner to master at his or her current stage of development (cf. Chapter 12.3.1). Skills that are within the learner's reach can be modeled by the coach (cf. Chapter 12.3.4), which involves demonstrating the skill (cf. Chapter 12.3.3) and providing explanations at each step that make explicit the actions being performed by the coach and their underlying reasoning processes. For example, the coach may take the transcriptions and sound charts compiled by the learner and show how to formulate hypotheses about the grouping of different sounds into functional units (phonemes) of the sound system of the target language on the basis of this data, explaining which sounds it is grouping together and according to what linguistic rules and principles it is determining the potential members of a phoneme. The learner watches these integrated performances and explanations and develops a conceptual model of the skill (cf. Chapter 2.2.3), which he or she applies in the following to solve problems of the same kind in the fieldwork environment. The coach monitors the learner during his or her problem-solving activities and provides assistance at appropriate moments to guide the learner toward a successful completion of the task, intervening when it detects the need or the opportunity for instruction or when the learner explicitly requests assistance (cf. Chapter 12.3.5). The coach's interventions may involve:

- Additional modeling (of the same or another required skill);
- Individualized feedback on the learner's decisions, results, and problem-solving strategies;
- Reminders, hints, and other information that help the learner to continue with the analysis;
- Instructional conversations (cf. Chapter 12.3.9) with the learner on the current problem to:
    - Identify and address misconceptions and gaps in the learner's task-relevant knowledge;
    - Guide the learner to a new perspective on the problem;
    - Demonstrate that the coach agent cares about the learner and his or her work;
    - Enkindle a genuine interest in the problem within the learner;
    - Encourage the learner to carry on with his or her work despite difficulties;
    - Increase the learner's awareness of linguistic phenomena;
    - Introduce the learner to new knowledge and skills;
    - Assess the learner's ability to solve problems in linguistic fieldwork.

Situations requiring the intervention of the coach typically arise when the learner's actions in the fieldwork environment do not yield the desired or expected outcomes (e.g. a particular analysis is not successful or produces a result that contradicts previous analyses and hypotheses of the learner). The coach does not try to prevent the learner from experiencing these kinds of failure because they lead to questions from the learner concerning what he or she did wrong and how he or she can do better (Schank & Neaman 2001: 49). These questions can be answered by the coach in different ways, for example by giving a (more or less specific) hint, asking a leading question, or sharing stories about how it has experienced and dealt with failures and successes in different fieldwork activities (p. 49, cf. Chapter 13.2.1). The latter is also an example of how the coach can provide encouragement when it detects or deduces a state of frustration in the learner.

Learners with different levels of proficiency in linguistic fieldwork require different degrees and kinds of support in their work. Beginning learners with little experience and much anxiety will require more assistance from the coach and more help with the basic skills involved in linguistic fieldwork than more advanced learners who have already mastered these skills. As learners are developing (and demonstrating) proficiency in linguistic field methods, the need for assistance from the coach agent declines, and the coach gradually fades from its initially prominent role into the background. However, it keeps monitoring the learner and is always ready to intervene, either taking the initiative or reacting to the learner's request.

The relationship between the learner and the coach agent is similar to the one between a human master and his or her apprentice. As indicated in the introductory quotation, this relationship is exclusive, between a single learner and his or her personal coach agent. Hence, unlike a human instructor, learners do not have to share their coach with other learners. Like the virtual native speaker, the learner has regular and extended interactions with the coach over a longer period of time. As a result, the coach has the opportunity to do its part in building a relationship with the learner (cf. Chapter 7.16). In contrast to the relationship between the learner and his or her peer agent (cf. Section 16.2.4), this is not a relationship between equals. The status of the coach as the master differs from that of the learner as its apprentice (although the learner as a human being is certainly superior to the coach as a computer program), which should be reflected in their interactions and mutual behavior. The coach should be portrayed as an individual that the learner perceives as more knowledgeable in linguistic fieldwork and capable of serving as a model and guide to help him or her to acquire the knowledge and skills of that 'trade.' As a master teacher, the coach is instructor-like and inspires respect and trust on the part of the learner (cf. Chapter 4.2). Most of the time, the interactions between the coach and the learner are serious and professional and focused on the learner's development in the area of linguistic fieldwork. The coach may occasionally use humor to lighten things up (cf. Chapter 10.3.2), but its role is not that of the "court jester" with a responsibility to entertain the learner. The learner may banter with a peer agent but not with his or her personal coach. While the relationship between the learner and the coach deepens over time, a certain respectful distance remains between the two. However, the coach's behavior to the learner should never be condescending. The respect shown by the learner should be returned by the coach.

16.2.4 The Peer

Individuals who are embedded in a group or community of learners can obtain information and support not only from their instructors and from educational material but also from their social interactions with other members of their "peer group." The term 'peer' broadly refers to an individual from the same group that learners perceive as having similar attributes, such as age, status, and level of knowledge or ability. Peers can provide learners with cooperation, competition, companionship, positive or negative examples, and the opportunity to teach others and be taught by them (cf. Chapter 12.3.10). The concept of the peer agent adds these kinds of social support to a linguistic fieldwork class.

The peer agent is basically in the same situation as the learner: confronted with the scenario of the virtual room, provided with access to the coach and the virtual native speaker, and equipped with the tools of the linguistic workbench, it explores the virtual room and tries to find out as much about the unknown language as possible. Like the learner, the peer agent has the status of an apprentice in linguistic fieldwork; hence, its knowledge of both the language under investigation and the linguistic methods necessary for its analysis is comparable to that of the learner, although the peer may know more (or less) about certain aspects of the language or a particular analysis method than the learner. However, in general, it does not possess expert-level knowledge of neither the language nor linguistic field methods.

Learning about the target language and about linguistic fieldwork is not the chief task of the peer agent, though. Its major responsibility is to change the learner's study of the linguistic fieldwork classes from a solitary into a social experience. (Both the virtual native speaker and the coach are working on this from their respective ends as well, but their primary roles are different.) The peer agent is equipped with skills in relationship building (cf. Chapter 7.16 and Chapter 10.3) and does its best to become acquainted with and form an amicable relationship with the learner. However, it does not force itself upon the learner and respects the sociocultural norms for approaching, addressing, and interrupting the learner (cf. Chapter 7.15).

In its social role, the peer agent is not meant to be a replacement for the learner's human peers (although it can serve as one if human company is not available). Instead, the peer is designed to blend into a hybrid community consisting of both human and artificial learners. This community is supported by a *collaborative learning and research environment*, to be added on top of the current platform of the Virtual Linguistics Campus, which provides:

- User awareness tools (Who else is online? Who is working on the same fieldwork (or other linguistic) problem? What have they done? What are their results?);
- Facilities for (a-) synchronous communication with one or several other (artificial or human) learners (chat, discussion forums, instant messaging, etc.);
- Support for bilateral and multilateral cooperation on linguistic problems, including work schedules, team meetings, milestones, etc.;
- Tools for collaborative authoring (e.g. weblogs and wikis, cf. Chapter 2.3.9);
- Means for pooling, organizing, storing, processing, and presenting results.

This collaborative environment provides the technical framework for different interactions between the learner and a peer agent in the linguistic fieldwork classes (cf. Chapter 12.3.10):

- *Cooperation*. The learner and the peer agent explore the virtual room together, dividing the work between them and pooling their results. For example, the peer collects utterances from the virtual native speaker, and the learner transcribes them. Or the learner explores one sentence type and the peer agent another, and afterwards they compare their analysis results. For assessment purposes, it is important that the contributions of the peer agent and the learner to an analysis are clearly distinguishable.
- *Competition*. The learner and the peer agent engage in a friendly contest about who manages to correctly describe a particular aspect of the target language (e.g. tense formation) first, who knows more words of the target language in a vocabulary test, who has more confirmed hypotheses in their portfolio, etc. The purpose of these competitions is to challenge the learner by setting him or her up against an opponent with comparable knowledge and skills, whom they have to outperform in their fieldwork activities in order to win.

- *Companionship*. The peer companion has no (current) fieldwork assignment of its own but takes an interest in the learner's current task and keeps the learner company while he or she is engaged in it. Being sensitive to the needs of the learner and the situation, the companion agent may compliment or encourage the learner, show empathy, or banter with the learner, but also offer ideas, opinions, and suggestions that may be useful to the learner in the context of his or her task (e.g. about how to classify a particular sound, what aspect of the virtual room to explore next, what previous mistakes of the peer to avoid, etc.). As a rule, the peer companion does not behave in a way that disrupts the learner's work or increases his or her extraneous cognitive load (cf. Chapter 2.2.2). When the learner makes it clear that he or she wants to be left alone, the companion does not bother him or her.
- *Troublemaker*. Similar to the companion, the troublemaker is with the learner during his or her fieldwork activities and likewise contributes its own views and suggestions about the learner's work. However, these contributions have to be taken with a pinch of salt because they are often incorrect or misleading (though not always). The challenge for the learner is to carefully evaluate the contributions of the troublemaker and to decide on their usefulness for his or her current work. The goal of pairing learners with an unreliable advisor is to put their knowledge and their confidence in their results and skills to the test again and again. It is also possible to give the learner both a companion and a troublemaker, which play the roles of a trustworthy and an untrustworthy advisor, respectively. These advisors both comment on the learner's current fieldwork activities. Often, the two will disagree and argue among themselves. Learners can follow these exchanges and obtain valuable insights for their own learning from the arguments provided by both sides.
- *Role model*. The peer agent performs a particular fieldwork task in a way that can serve as a model of best practices in linguistic fieldwork for other learners. The peer agent presents its work (cf. Chapter 12.3.12) in a public forum, making explicit the steps in its approach and the reasoning behind them. It is also possible for hybrid teams of learners and peer agents to become role models and give these presentations together. In general, peer agents as role models will only work if the learning culture in the linguistic fieldwork environment values cooperation higher than competition. It should be noted that a role model does not have to be positive. That is, the peer agent can also give a bad example illustrating how linguistic fieldwork should *not* be done. The performance of the negative role model can be dissected, for example by the coach agent or by the learner, to identify suboptimal practices which the learner should avoid.
- *Peer coaching*. Advanced learners can coach a less knowledgeable peer agent on certain aspects of linguistic fieldwork, for example on how to analyze an utterance of the virtual native speaker into its sentence- and word-level components or how to extrapolate from the collected linguistic data to compose previously unseen forms of the target language. By serving as a model and guide for someone else, learners can deepen their own understanding of the subject matter.[262] Conversely, peer-coach agents can also share their greater knowledge and skill in particular areas with the learner, for example by critiquing his or her

[262] The role of the peer 'coachee' is similar to the one of the *teachable agent (TA)* (Biswas et al. 2005). Learners teach this kind of pedagogical agent and then ask it questions or have it solve problems. The TA generates answers and solutions, making its reasoning visible in a graph structure called a *concept map*. Learners review the TA's results and see where they need to revise its knowledge (and their own) (Blair et al. 2007).

analyses or offering hints and other information. Since a peer coach is perceived to be more like the learner than, say, the coach agent, the learner may be more open to its suggestions.

In the course of a fieldwork class, the learner may meet a variety of peer agents with different attributes and qualities. With some of these agents, the learner may form a closer relationship than with others, in particular with those which he or she perceives to be like him or her with respect to their interests and personality, according to the principle of similarity attraction (cf. Chapter 9.2).[263] In some units, the learner may interact with a single peer or with a small group of peer agents, whereas in other units, he or she may prefer to work on his or her own. Depending on the learner's needs and preferences, peer agents can engage the learner in different interactions from among the set described above. However, care should be taken when casting a given peer agent for several of these subroles, to avoid inconsistent role assignments (cf. Chapter 7.14) across a series of fieldwork sessions. For example, the learner is unlikely to trust a peer agent that played the role of troublemaker in one session when it wants to be the learner's companion in a later session.

16.2.5 Summary of the Roles

In this section, three roles for embodied interactive pedagogical software agents were described which provide learners with different kinds of support while they are studying a linguistic fieldwork class on the Virtual Linguistics Campus. The virtual native speaker produces utterances in its native language, which are analyzed by the learner in order to formulate hypotheses about that language. The role of the virtual native speaker has considerable (currently neglected) potential with respect to interaction and relationship building. Section 16.2.2 described an enhanced virtual native speaker agent that can interact more flexibly with learners and works to establish a productive working relationship with them.

Whereas the virtual native speaker is primarily responsible for providing learners with the raw linguistic data for his or her analysis, the coach (cf. Section 16.2.3) serves as a master teacher helping its apprentice (the learner) to develop the linguistic skills necessary to analyze that data and to form hypotheses about the target language based on his or her analysis. The coach provides modeling and guidance for the learner and gradually fades into the background as the learner becomes increasingly able to accomplish the fieldwork tasks on his or her own.

The peer agent (cf. Section 16.2.4) adds a third element to the linguistic fieldwork classes: social learning with, through, and from other members of the same group in the context of a learning community, which are comparable to the learner in terms of status, situation, and level of ability. Peer agents can be designed to cooperate or compete with the learner in the virtual room, keep him or her company during fieldwork sessions, challenge his or her skills and confidence, serve as models of best (or worst) fieldwork practices, and provide or receive peer coaching on linguistic field methods.

Each of these three roles is associated with its own set of capabilities, expectations, and responsibilities. Pedagogical agents that are designed for them not only need to be equipped to function effectively in their appointed role but also have to convincingly convey an appealing

[263] Kim (2007) found that a learning-companion agent with characteristics similar to those of the learner positively influenced the learner's cognitive and affective accomplishments. A learning companion with about the same level of competence as the learner was perceived more favorably and led to higher self-efficacy and recall.

agent persona that is appropriate for the role and should do their best to develop a relationship with the learner (cf. Chapter 8.4). As the previous chapters have shown, to achieve the triple goal of function, persona, and relationship building, many different considerations must enter into the design of such an agent, which are illustrated in the following section using the role of the virtual native speaker as an example.

16.3 The Design of the Virtual Native Speaker

The linguistic fieldwork classes already feature an embodied agent: the virtual native speaker. However, as pointed out in the previous section, this agent currently underachieves with respect to its potential for interaction and relationship building. This section offers detailed suggestions for enhancements to the virtual native speaker that promote it from an automaton that plays back canned utterances on demand to a more flexible, competent, and interesting partner for learners. The enhancements are discussed in separate sections corresponding to the agent design components discussed in the previous chapters of the thesis: embodiment (cf. Chapter 9), behavior (cf. Chapter 10), conversation (cf. Chapter 11), expertise (cf. Chapter 12), emotion and personality (cf. Chapter 13), architectures and platforms (cf. Chapter 14), and evaluation (cf. Chapter 15).

16.3.1 Embodiment

In the current version of the linguistic fieldwork classes, the virtual native speaker is already implemented as an embodied character with rudimentary interactive capabilities. This section presents design considerations for the embodiment of the enhanced virtual native speaker agent described in Section 16.2.2.

As discussed in Chapter 9.1, agent designers have to decide what kind of entity the body of their agent should represent. Given that the virtual native speaker is intended to portray a speaker of a natural human language, non-human embodiments, including animals, fictitious entities, objects, and shapes, do not qualify as personifications of the agent. Hence, the virtual native speaker is appropriately represented as a virtual human. However, portraying the virtual native speaker (or the coach or the peer agent, cf. Section 16.2.3 and Section 16.2.4) in the image of human beings involves the danger that learners might overestimate the agent's capabilities. This danger increases as human-like conversational and other capabilities are added to the agent (cf. Section 16.3.3 and Section 16.3.4). In order not to mislead the learner with its human embodiment, the agent should exhibit cues that clearly indicate its role as the learner's informant in the fieldwork setting and thus guide the learner's expectations about its capabilities and limitations (cf. Chapter 7.14). These cues are provided by the agent's appearance, behavior, and interactions. For example, if the agent looks like a speaker of a foreign language, greets the learner in that language, or is explicitly introduced as the learner's informant, the agent's role in the virtual room and the range of interactions that are possible become clearer to the learner.

While the current virtual native speaker is essentially a 'front-end' for the human native speaker who provided its inventory of utterances, the enhanced speaker agent is a separate individual from the culture where the language of the fieldwork class is spoken, with its own identity, personality (cf. Section 16.3.5.2), back-story (cf. Section 16.3.5.3), and idiosyncrasies (cf. Section 16.3.2.2). The appearance of the virtual native speaker has different functions:

- It authentically portrays a contemporary native speaker of the target language and culture.
- It manifests the agent's status as a unique individual (cf. Chapter 9.3).
- It helps the learner to quickly understand the speaker's identity and its role as informant.
- It attracts the learner and invites him or her to interact with the virtual native speaker.

Age and gender of the virtual native speaker are currently determined by the human voice talent providing the speaker's utterances; however, given access to a speech synthesis system with sufficient intelligibility, naturalness, and expressiveness (cf. Chapter 10.1.1 and Chapter 11.1.2), it may be possible to overcome this restriction (the speaker's voice is discussed in Section 16.3.2.1.1 below). The speaker agent portrays the ethnicity of a member of the culture in which the target language is spoken as the first language. The ethnicity manifests itself in the primary (physical) attributes of the agent, including its facial features, skin and hair color, physique, etc. but also in its secondary attributes, such as clothing, hair style, accessories, and so on. (Of course, the agent's portrayed age and gender also play a role in shaping these aspects.) Designers may consider exploiting existing stereotypes about the appearance of members of the culture represented by the speaker (consider, for example, Reiko's kimono and traditional hair style in Figure 94), but they should research the cultural background first to find out to what extent these stereotypes apply to the speaker they intend to portray. Currently, the majority of virtual native speakers in the VLC fieldwork classes represent younger people (due to the relatively young age of the human voice talents), who may rarely dress in the traditional attire of their culture (hence, the appearance of the virtual native speaker Reiko is probably inappropriate for the younger Japanese female it portrays). In general, the speaker agent's appearance should be as authentic as possible with respect to the contemporary fashion and conventions of the target culture. However, authenticity does not imply uniformity. The virtual native speaker is an individual and as such has its own style, not least with respect to its appearance. Therefore, it should be designed to stand out from other actors in the virtual room in terms of hair style, clothing, accessories, bodily features, etc. In particular, learners' attention will be attracted by attributes of the agent that appear unusual or unfamiliar to them, such as a tattoo, a feather in the hair, or brightly colored clothes. To support the impression that the virtual native speaker is a character with a life of its own even outside the fieldwork sessions, all these aspects should exhibit some variation in the course of a fieldwork class. For example, no person in the real world wears the same clothes for weeks in a row, so the virtual native speaker should also appear in different clothes in different fieldwork sessions. Similar to a real person, the speaker agent may have a favorite set of clothes or (occasionally) new clothes which it shows to the learner and other virtual actors.

To appeal to the learners of a linguistic fieldwork class, the virtual native speaker should be made physically attractive for its target audience (cf. Chapter 6.4 and Chapter 9.2), however without sacrificing its sociocultural identity by trying to model the agent too closely on the learners' concept of ideal beauty (which is subjective and from a different culture). The speaker needs physical attributes that are considered attractive across cultures, such as health, symmetry of face and body, and a straight profile. Hair style, cosmetics, costume, and accessories are further important visual clues of attractiveness, but these attributes are subject to individual taste and cultural acceptability (both from the learner's and the designer's perspective). In addition, their importance for expressing the agent's cultural identity imposes restrictions on their design. In general, a culturally authentic virtual native speaker will exhibit at least some characteristics that appeal to members of its own culture, although not necessarily to members of other cultures, including the learner.

Movement (cf. Chapter 9.4) is a central capability of the enhanced virtual native speaker's embodiment. The speaker agent uses movement of the body and body parts to communicate messages, act out scenes in the virtual room, and interact with the learner. Unlike the current, essentially immobile virtual native speaker, which always sits behind a table and therefore only requires a head, arms, and a torso, the enhanced speaker agent possesses a complete, animated body capable of expression, gesture, and locomotion, which it uses during its performances with other actors in the virtual room and during its interactions with the learner, as described in Section 16.3.2. To create the necessary animations for the virtual native speaker (and for the coach and peer agents as well), behavior sequencing (cf. Chapter 14.2.1) is the most straightforward approach given the current implementation of the virtual room and the capabilities of Adobe Flash ([INT 161]), the animation technology used to create the virtual room, the current virtual native speaker, and the other animations in the Virtual Linguistics Campus. The agent's complex animations are composed by dynamically sequencing primitive animations from a pre-designed behavior space at run-time. The creation of this behavior space will require considerable work for each agent because the animations have to be more complex, individual, varied, and smoother than in the current implementation of the virtual native speaker. A behavior generation engine accepting motor commands from the agent's brain and rendering the corresponding animations, as in the layered generation approach (cf. Chapter 14.2.2), would require less work from animators and provide more flexibility with respect to handling interruptions of behavior sequences (for example by actions of the learner or another virtual actor). However, such a behavior generation engine may be costly to develop and difficult to integrate into the environment of the virtual room because it would have to be implemented as an additional plug-in component for the web browser to render the animations with sufficient speed. An enhancement to the behavior sequencing approach, which allows the combination of planned with unplanned agent animations, for example in response to the learner's actions, as in state machine compilation (cf. Chapter 14.2.3), may be more feasible. While the (enhanced) behavior sequencing approach is sufficient for generating the animations of a single agent, performances involving interactions between multiple characters in the virtual room require a multi-character platform (cf. Chapter 14.2.4), which controls the individual characters and orchestrates the performance.

Embodied agents can be rendered with different degrees of realism and depth of the images portraying the agent (cf. Chapter 9.5). For the virtual native speaker and the other agents in the linguistic fieldwork classes, a 2½D artistic representation will be adopted, similar to the way the virtual native speaker is currently represented (cf. Figure 94). There are several reasons for this decision:

- The agent inhabits the virtual room and should therefore be created with the same depth and level of realism. A photorealistic, three-dimensional character would look out of place in a hand-drawn environment with smaller depth.
- The camera position in the virtual room is fixed. Learners do not have the possibility to view the room and the virtual native speaker from different angles. Otherwise, a 3D model of the room and the speaker agent would be required.
- 2½D (i.e. slightly deeper than 2D) is necessary because the virtual room has a certain depth and the agent should have the ability to move around and perform within the depths of the room without looking 'flat' in comparison.
- By choosing an artistic rather than a realistic representation, the pitfall of the uncanny valley can be avoided (i.e. of creating a character that is almost human but still falls short and hence elicits strong negative responses from learners, cf. Chapter 9.5). Furthermore, a lower level

of realism may reduce learners' expectations of human-like behavior and capabilities, which allows the design to compensate for the limitations of state-of-the-art machine intelligence.

- A 2½D animated artistic embodiment can be easily created with Adobe Flash, using a behavior sequencing approach (see above). In general, representations below 3D are easier to draw and animate while permitting greater levels of detail, and require less processing (cf. Chapter 9.5).

In the current implementation, the speaker has a partial human embodiment consisting of the upper half of the body. While this is sufficient as long as the speaker only has to deliver spoken utterances, a complete human body with a face, hands, feet, and a torso (cf. Chapter 9.1) allows the virtual native speaker to leave its current place behind the table (cf. Figure 94), move around the virtual room, and participate fully in interactions with the learner and other virtual actors.

The embodiment of the virtual native speaker plays an important role in helping learners to locate their informant in the virtual room and to recognize the speaker agent as a virtual person. Hence, it is not appropriate for the virtual native speaker to switch to a different embodiment in the course of the fieldwork class or even during interactions with the learner. However, certain aspects of the agent's appearance may change to enhance its believability as an individual (see above); still, these changes should never alter the identity of the virtual native speaker.

16.3.2 Behavior

As an actor in the virtual room of a fieldwork class, the virtual native speaker makes use of the features of its embodiment to perform a variety of behaviors, which serve to interact with the learner and other virtual actors, to enhance its own believability as a persona, and to build a relationship with the learner. This section discusses the design of communicative, believability-enhancing, and relationship-building behaviors for the virtual native speaker.

16.3.2.1 Communicative Behaviors

The communicative behaviors of the virtual native speaker can be subdivided into verbal and non-verbal behaviors. The design of the former involves considering aspects of the speaker's voice, whereas developing the latter is concerned with how the agent uses facial expressions, gaze, gesture, posture, and locomotion in its interactions with others.

16.3.2.1.1 Voice

The voice is a central property of the virtual native speaker, which the agent uses to provide linguistic data to the learner for analysis and to distinguish itself from other actors in the virtual room. The voice conveys the speaker's utterances that contain the raw data which the learner has to analyze in order to write the grammar of the virtual native speaker's language. Therefore, the articulation of the virtual native speaker has to be authentic and intelligible, accurately modeling the segmental and suprasegmental aspects of the language, i.e. its inventory of speech sounds (phonemes) and its prosody (as described by patterns of tone, intonation, rhythm, and stress), as well as allowing learners to identify higher-level units and constructions (word, phrasal, and sentence patterns). However, the speaker's voice also has to be natural and expressive. It should not be tiresome for the learner to listen to the virtual native speaker, even if its utterances are longer. Furthermore, the voice is a vehicle for expressing the speaker's current emotional and cognitive state in its interactions with the learner and other virtual actors. Finally, the voice is a unique identifying characteristic of the virtual native speaker and hence should differ from those of other speakers of the same or a different language in the virtual room and the linguistic fieldwork classes.

The speaker's voice is currently implemented as recordings of a human native speaker of the language. While recorded human speech satisfies the criteria of authenticity, intelligibility, and naturalness, the current virtual native speaker is unable to generate its own utterances in order to describe new situations in the virtual room or to participate in conversations with other virtual actors or with the learner (cf. Section 16.3.3). For flexible utterance generation, the virtual native speaker has to use synthetic speech (cf. Chapter 10.1.1 and Chapter 11.1.2). The intelligibility of synthetic speech at the segmental level is no longer regarded as a problem (at least when the task is to synthesize a sequence of words that can be correctly recognized by speakers of the language; the situation might be different when the listener does not know the language and has to transcribe the utterances, as in the virtual fieldwork scenario). However, even the best concatenative speech synthesizers (cf. Chapter 11.1.2.2) are still inferior to a human voice in terms of naturalness and expressiveness, which becomes apparent when utterances are longer and/or consist of multiple sentences. The artificial-sounding prosody of synthetic voices is a problem for fieldwork tasks concerned with prosodic aspects of the target language (e.g. intonation and stress patterns). Still, the use of synthetic speech may be feasible for generating utterances that illustrate other aspects of the language than suprasegmental phenomena. The deficiencies of current speech synthesizers are the least perceptible in short utterances with simple prosodic patterns, which is a restriction that is often acceptable in the linguistic fieldwork setting because it limits the amount of linguistic data and the range of phenomena presented to the learner in a single utterance. (Currently, the recorded utterances of the virtual native speaker generally also consist of a single word, phrase, or sentence.)

Concerning the expressiveness of the virtual native speaker's voice, there are problems with using either human or machine-generated speech. A human voice can express a much wider range of different emotional and cognitive states than a synthetic voice, but only in pre-recorded utterances. In contrast, synthetic speech can convey all kinds of dynamically generated messages, but creating the appropriate emotional (or other) tone for a synthetic utterance is difficult, although some experiments have successfully manipulated parameters of speech synthesizers to express different emotions in synthetic speech (cf. Chapter 13.4.2). The default prosody of many commercially available speech synthesis systems is neutral and typically not adaptive to the emotional tone of the message. However, it may be possible to compensate for the lack of expressiveness in the synthetic voice by relying more on non-verbal behaviors, such as facial expressions, to express the speaker's emotions and other internal states.

Casting the appropriate voice for the virtual native speaker is a problem that should not be underestimated because the voice chosen does not only affect the speaker's ability to perform its primary task (i.e. to provide raw linguistic data from its native language) but also the speaker's image as an individual and as a representative of the target culture. As an individual, the virtual native speaker has its own idiosyncratic ways of conveying its internal states through its voice and other means (cf. Chapter 13.4.2), which contribute to manifesting the speaker's identity, as does the accent of the speaker's voice, which indicates the place of origin of the individual portrayed (cf. Chapter 10.1.1). Several virtual native speakers of the same language may have different accents, introducing the learner to different varieties of the target language. The cultural background of the virtual native speaker establishes rules and conventions that determine what the agent expresses how, when, and to whom, using its voice and other means (cf. Section 16.3.5.1.3). By observing the agent's ways of expressing itself to others, the learner can identify these rules and conventions and compare them to those of his or her own language and culture. The design of the virtual native speaker's voice should acknowledge both these cultural and individual influences. Again, this is easier to realize with a human voice talent than with a speech synthesizer. However, there are already speech synthesis systems that provide voices for different varieties of a language (e.g. American vs. British English).

Whether a human or a synthetic voice is chosen for the virtual native speaker, it should be consistent with the other aspects of the agent, such as its embodiment and personality (cf. Chapter 7.5). For example, a male voice is clearly inappropriate for a female speaker, as is a soft and quiet voice for an extrovert character. Furthermore, the level of realism adopted for the virtual native speaker (artistic rather than photorealistic, cf. Section 16.3.1) influences the casting of the voice. As discussed in Chapter 10.1.3, people have been found in experimental studies to prefer consistent combinations of face and voice that do not mix human and synthetic faces and voices. Hence, the combination of an artistic face with recorded human speech (as in the current speaker) might be perceived as inconsistent, and a synthetic voice might be more appropriate for the virtual native speaker from the point of view of consistency (cf. Chapter 10.1.3). It is difficult to reconcile the conflict between the requirements of consistency and speech quality because of the still inferior capabilities of speech synthesizers. However, the negative effects of using an inconsistent combination of artistic character and human voice may not be that severe given that this approach has been used in animated films to the present date and learners may be familiar with it.[264]

Each fieldwork class has its own virtual native speaker, so a different voice has to be cast for each of them. Furthermore, a fieldwork class may feature several speaker agents that participate in multi-character performances including spoken conversations in their native language, which means that different voices for the same language are required. It can be difficult to find the necessary number of human voice talents to give each virtual actor a distinct voice. Synthetic speech has a similar problem: a number of commercial systems provide several voices for the same language, but typically only for the more 'important' languages, such as English and German. It is always possible to build new synthetic voices, but the time, effort, and cost involved may be considerable (cf. Chapter 11.1.2.2).

Given that both recorded human and synthetic speech have their respective strengths and weaknesses, designers might consider using both for the virtual native speaker. However, a hybrid approach mixing synthetic and natural human speech is not acceptable for the virtual native speaker (or the other agents in a fieldwork class) because such a combination would be perceived as inconsistent by listeners (cf. Chapter 10.1.1). Developers should use either recorded human speech or synthetic speech for the virtual native speaker but refrain from mixing the two.

If synthetic speech is considered for the virtual native speaker, a high-quality concatenative speech synthesis system should be chosen. Whereas a concept-to-speech approach (cf. Chapter 11.1.2.3) allows a more seamless integration of speech synthesis with the process of message generation (cf. Chapter 11.1.4), such systems are still much rarer on the market than text-to-speech synthesis systems. Since the resources for developing a proprietary solution are not available in the VLC team, existing text-to-speech products have to be acquired and integrated

[264] Having said that, there is an important difference between the virtual native speaker and characters in animated films: the learner interacts with the former while he or she only watches the performance of the latter. The effects of interaction on the importance of perceived consistency in virtual interaction partners are not clear yet and require more research.

with the other components of the virtual native speaker (cf. Chapter 7.11). While speech synthesizers have already been developed for many languages, including German, English, French, Dutch, Spanish, Italian, Portuguese, Swedish, Norwegian, Finnish, Estonian, Icelandic, Czech, Russian, Greek, Croatian, Turkish, Romanian, Breton, Japanese, Chinese, Korean, Hebrew, Arabic, and Indonesian ([INT 215]), less common languages can be a problem because no one has built a speech synthesizer for them (yet). In addition, the available systems for the same and for different languages do not exhibit the same level of output quality, which can make the process of finding the right voice difficult.

16.3.2.1.2 Gestures

In the current linguistic fieldwork classes, the virtual native speaker relies primarily on its voice to participate in interactions with the learner. In contrast, non-verbal communicative behaviors are generally underused by the actors in the virtual room, including both the virtual native speaker and its supporting actor (that appears in later units of a fieldwork class). However, as argued in Chapter 10.1, non-verbal communicative behaviors have important functions in regulating conversation, by contributing content to the ongoing interaction and conveying emotional and cognitive states of the speaker. Hence, gestures, facial expressions, gaze patterns, posture, and locomotion will be added to the behavioral repertoire of the enhanced virtual native speaker and the actors supporting it. Gestures are discussed below, whereas the other non-verbal communicative behaviors are covered in the following sections.

The virtual native speaker performs gestures during its interactions with the learner and with other virtual actors. The range of gestures occurring in the speaker's behavior includes not only the conscious emblematic and propositional gestures (where appropriate in the context of the interaction) but, importantly, also the iconic, metaphoric, deictic, and beat gestures that are performed spontaneously during the speaker's utterances, as described in Chapter 10.1.2. The virtual native speaker requires a sophisticated model of face-to-face communication, which integrates the agent's speech with its gestures (and other non-verbal behaviors) in ways that emulate the human use of non-verbal means in face-to-face conversations. One important requirement for the speaker's multimodal communication is deictic believability. Since the virtual native speaker constantly refers to entities at different locations in its environment, its utterances should make use of natural and unambiguous combinations of speech, gesture, and body movement (cf. Chapter 4.2). For each embodied referential utterance, the appropriate verbal and non-verbal elements have to be identified, synchronized, and assembled into a coherent behavioral sequence.

Similar to the speaker's voice, its non-verbal behaviors not only serve to communicate messages but also reveal important things about the speaker as an individual. The manner in which the virtual native speaker performs gestures and other behaviors should be appropriate for its personality (e.g. an introvert speaker should not display emphatic gestures, cf. Section 16.3.5.2), its cultural background, and its current internal state (e.g. an agent that is confused, tired, or sad may perform the same gestures differently from one that is unperplexed, rested, or happy). In addition, the speaker's behaviors should not come "off the shelf." While it may be tempting to design a single set of gestures (or other behaviors) and reuse them for all virtual native speakers, this approach does not help to create characters with individuality (cf. Chapter 7.8). Instead, each virtual native speaker should add a personal touch to its performance of different communicative behaviors. Furthermore, the same behavior should be performed with some degree of variability (cf. Chapter 7.9) across different occasions. In a behavior sequencing approach (cf. Chapter 14.2.1 and Section 16.3.1), these requirements imply a considerable workload for animators because the behavior space has to be populated with a much larger number of primitive animation elements. In contrast, both individuality and variability are much easier to achieve with a parameterized behavior generation engine.

The virtual native speaker is designed to represent a different culture; hence, not only its appearance (cf. Section 16.3.1) but also its behaviors should be authentic with respect to that culture. The speaker agent should portray gestures, facial expressions, etc. in the way they are displayed in its own culture, not the culture of its designers or the culture of the learner. Since both the inventory of gestures and the ways of performing them are subject to cross-cultural variation, the cultural background of each virtual native speaker requires careful consideration before its gestures are designed. When there is a possibility for the speaker's gestures to cause misunderstandings or offence, as in the case of some emblematic gestures (cf. Chapter 10.1.2), additional information should be provided that clarifies their meaning.

16.3.2.1.3 Face

Each virtual native speaker is a different individual, which is most immediately noticeable in the face with which it presents itself to the learner. The face manifests the speaker's identity and cultural affiliation and portrays different kinds of expressions that convey the agent's current state and social intent (cf. Chapter 10.1.3). Through direction of eye gaze, the face indicates the speaker's current focus of attention. Furthermore, the speaker's lip movements are synchronized with its verbal utterances.

The face of a virtual native speaker is an important element by which it distinguishes itself from other speakers of the same language or different languages. When several virtual native speakers appear in a fieldwork class, each should have its own unique face to allow the learner to recognize and remember the speaker. The face should also identify the virtual native speaker as a member of its culture but without giving the impression of being stereotypical. That is, the 'typical' facial features (eye shape, skin color, etc.) of the group of people represented by the speaker should be present but not overemphasized in the design of the face.

The screen area occupied by the face of a fully embodied agent, such as the virtual native speaker, is small in comparison to the space available for the face of a talking-head or face-only agent. Still, the face of the virtual native speaker should exhibit sufficiently detailed features to convey its identity and to communicate different emotional, cognitive, and physical states. A photorealistic face is not required, but for the generation of facial expressions, the eyes, mouth, and eyebrows are essential components, which the speaker agent should be able to transform into different shapes and combine into different facial displays (cf. Chapter 10.1.3) that complement its utterances.

While speaking, the auditory signal is synchronized to observable articulatory movements of the lower part of the face (i.e. the lips, tongue, teeth, and jaw). Audiovisual speech is natural human behavior and therefore also expected of the virtual native speaker. In fact, it would be very unusual if the virtual native speaker delivered its utterances without any visible articulation. So the current speaker already performs lip movements that accompany its utterances. However, the speaker's visible articulation is a façade rather than an accurate display: neither are the different articulatory facial movements completely and faithfully implemented nor are they accurately synchronized to the auditory speech. In essence, the visual articulatory displays of the current virtual native speaker are a gross simplification compared to a real human speaker. While implementing a more faithful visual articulation and synchronizing it to the character's speech would mean significantly more work, there are also a number of benefits. First, it would improve the intelligibility of the virtual native speaker's articulation, due to the complementary nature of auditory and visual information in speech (cf. Chapter 10.1.3). Second, facial animation synchronized with speech can provide accurate displays of the articulation of speech sounds, which are instructive for learners. A function for zooming in on the virtual native speaker's face could be added, which allows learners to focus on the speaker's articulation. Unlike a human speaker, the skin of the virtual native speaker could become transparent and reveal the motion of the articulators. By observing the visual articulation, learners may obtain information that helps them to clarify what sounds they heard. Of course, the articulatory apparatus and movements of the enlarged face would have to be detailed and accurate for this purpose. Third, visible articulation that is only slightly out of sync with auditory speech disturbs human observers and therefore significantly reduces the speaker's believability as an individual. This effect can also be observed in video clips of human speakers in which the audio lags behind the video: viewers' comments on these clips are typically negative and mention the temporal inconsistency between the auditory and the visual signal. In short, people expect consistency between auditory and visual speech, which has to be taken into account when designing these aspects of the virtual native speaker. A similar preference for consistency between face and voice was discussed in Section 16.3.2.1.1 above and in Chapter 10.1.3 with respect to different pairings of human and synthetic faces and voices.

16.3.2.1.4 Gaze

While eye gaze (the direction where an individual is looking with his or her eyes) has important functions in face-to-face communication (cf. Chapter 10.1.4), it serves none of these functions in the interactions of the current virtual native speaker, which does not exhibit any variation in its gaze patterns but rather always looks in the same direction, i.e. at the learner (although it may display some random eye movements). In contrast, the enhanced virtual native speaker varies the direction of its gaze to clarify its verbal messages, send social signals to other participants in the interaction, open channels to them in order to obtain information, manage the flow of the dialogue with the learner, and, finally, communicate social and individual emotions. These communicative functions of eye gaze were discussed in detail in Chapter 10.1.4. To give a simple example of how the use of eye gaze can improve the interaction between the learner and the virtual native speaker and enhance the speaker's believability, when the learner clicks on an object in the virtual room, the current speaker instantly provides the corresponding word, looking at the learner instead of the object referred to, even if the object is behind the speaker's back. In contrast, a human speaker would have to turn his or her head and/or body and gaze at the object of interest first in order to understand the target of the learner's pointing gesture. By using gaze and head/body movement like humans do, the virtual native speaker can establish a common ground with the learner, confirming that it is referring to the same object as the learner with its utterance (Rickel et al. 1999). Furthermore, the speaker becomes more believable as an individual. In general, it is quite unnatural behavior for a human being to always look in the same direction, whether within or outside interactions with others. Instead, the enhanced virtual native speaker may choose to direct its attention toward different objects or events in the virtual room that draw its interest, even without being prompted by an action of the learner. Of course, the speaker's gaze patterns should indicate that it is paying attention to the learner, but without giving the impression of monitoring him or her. The learner might feel uncomfortable if he or she is being watched all the time (cf. Section 16.3.2.3).

When portraying eye gaze, the virtual native speaker should adhere to the rules of its culture for using gaze in communication. Speakers representing contact vs. non-contact cultures should be designed to exhibit different patterns with respect to the direction, duration, and frequency of gaze (see the examples provided in Chapter 10.1.4). As with gestures, it may be necessary to provide additional information that explains differences between the gaze behavior of the virtual native speaker and the learner.

16.3.2.1.5 Posture and Locomotion

The human embodiment constrains the range of behaviors that the virtual native speaker is expected to be able to perform, which also includes the possible positions or arrangements of its body and limbs while not moving (i.e. its different postures, cf. Chapter 10.1.5) and the ways in which it can transfer itself from one location to another (i.e. locomotion, cf. Chapter 10.1.6). The current virtual native speaker only assumes a single posture: sitting upright behind a table in the virtual room (cf. Figure 94). There is neither variation in the speaker's posture nor a change of location during a fieldwork session. Due to the lack of a complete body, the speaker cannot leave its place behind the table, i.e. it is incapable of locomotion. In contrast, the enhanced virtual native speaker has the ability to portray different postures and variants of the same posture, as well as to leave its default location behind the table in order to participate in interactions taking place at a different place in the virtual room (e.g. talking to another virtual actor at the door or at the window) and to move toward objects of interest in order to be able to manipulate or make unambiguous reference to them (e.g. walking toward the lamp and turning it on (or off) or picking up the glasses on the sideboard and putting them on while providing the appropriate utterance, cf. Figure 94). Locomotion not only involves crossing a distance but also moving around certain obstacles that may appear in the speaker's path.

Given the constraints imposed by the environment of the virtual room, it is not necessary to give the virtual native speaker the full range of human forms of locomotion (cf. Chapter 10.1.6). For example, swimming is not possible in the virtual room. Other locomotive behaviors are restricted: there is not much room for running around, and even fewer opportunities exist for climbing, for instance. The virtual native speaker's default mode of locomotion is walking. Concerning the range of postures for the virtual native speaker, the most common ones will be sitting and standing (cf. Chapter 10.1.5).

Similar to the other behaviors, posture and locomotion should exhibit both individuality (cf. Chapter 7.8) and variability (cf. Chapter 7.9). Furthermore, posture and locomotion should provide cues as to the speaker's personality, current state, cultural origin, and relationship to the learner and other virtual actors. For example, the speaker may show an upright posture and bold movements to portray a self-confident, extrovert individual. In contrast, a drooping posture and slow movements may signal fatigue or sadness. Japanese women (like Reiko in Figure 94) are said to be traditionally "pigeon toed," i.e. to walk with their toes turned inward (although this may no longer be true for many modern women in Japan). A forward-learning posture can show that the speaker likes the learner (cf. Chapter 10.1.5). Since posture is also an indicator of physical attractiveness, the virtual native speaker should be designed with a well-balanced or erect posture, which is almost universally regarded as more appealing than a misshapen stooping figure.

The virtual native speaker need not be in the same place every time a learner enters the virtual room. Instead, it could be busy dusting off the shelf in the back, looking out of the window, or even coming into the room through the door at the same time as the learner. However, as soon as the learner is there, the speaker should stop doing whatever it is doing and make itself available as the learner's informant (a different behavior would be interpreted as uncooperative). The place behind the table may continue to serve as the default location of the enhanced virtual native speaker, providing a known and reliable place for the learner to obtain linguistic data from the target language and functioning as the starting and end point for the speaker's forays into the virtual room in the course of its interactions with the learner and with other virtual actors.

16.3.2.2 Believability-Enhancing Behaviors

Apart from lip movements that accompany its utterances, the only non-verbal behaviors that the virtual native speaker currently performs include a small set of animations that are intended to enhance the speaker's believability as an individual, such as eye-blinking, smiling, sighing, yawning, pursing the lips, and small arm and chest movements. The speaker performs these behaviors when it is idle, i.e. waiting for the learner's next action. As the set of idle behaviors is small (most virtual native speakers have only two or three of them) and the behaviors are performed randomly, they quickly become repetitive and contextually inappropriate, destroying the illusion of life that they are intended to create. Furthermore, the idle behaviors sometimes draw unnecessary attention to the virtual native speaker and may even give the learner some incorrect ideas about the speaker. For example, the virtual native speaker Reiko performs one idle behavior (yawning) that could be interpreted by the learner as suggesting boredom and lack of interest, which is not conducive to the process of building a relationship with the learner.

The design of the enhanced virtual native speaker follows a different philosophy, which de-emphasizes the role of explicit believability-enhancing behaviors and attempts to achieve believability more indirectly through verbal and non-verbal communicative behaviors that observers perceive as human-like, individual, and culturally authentic. The agent also performs behaviors to enhance its persona, but these behaviors are more subdued in comparison to the current implementation, providing the background for the speaker's interactions and performances.

Since the virtual native speaker portrays a human being, it shows a number of reflexive behaviors that are required by human physiology, such as breathing, eye-blinking, and wetting the lips (cf. Chapter 10.2). While they do not serve a real need, the speaker nevertheless performs these automatic behaviors with their natural regularity because they are expected of a human-like entity. (For example, it would be unnerving for learners if the speaker was always staring (at them) without ever blinking.) Furthermore, the virtual native speaker may possess idiosyncrasies, even quirks, to support its character (cf. Chapter 7.8). For example, it may have the habit of scratching its chin when thinking, playing with objects on the table when bored or nervous, shaking its hair out of its eyes when they get in the way, or looking over its glasses when they have slid down. The conditions listed for each example of idiosyncratic behavior indicate that these behaviors are performed in a context-sensitive fashion rather than randomly, and in a way that is prominent enough to be noticed by the learner but not eye-catching enough to be a distraction. As a general rule, no idiosyncratic behavior of the virtual native speaker should interfere with its primary role as the learner's informant on the target language, by drawing the learner's attention to a behavior that is irrelevant for analyzing the speaker's current utterance. In addition, all believability-enhancing behaviors should exhibit sufficient variability (cf. Chapter 7.9) to avoid the impression of repetition.

16.3.2.3 High vs. Low Profile

As the learner's informant, the virtual native speaker is essential for the learner to succeed in his or her fieldwork tasks. Therefore, it is present and available every time the learner visits the virtual room, providing a familiar place where the learner can obtain the linguistic data for his or her analysis of the unknown language. The virtual native speaker has to be constantly available, which means that it cannot simply disappear when its services are not required. It should always be clear to the learner where he or she can find the speaker in the virtual room when in need of further linguistic input. However, constant availability does not imply that the virtual native speaker just sits there all the time and patiently waits for the learner to ask about things or events in the virtual room, as in its current implementation. This is unnatural behavior (no human informant would have the patience or the time to do that). Furthermore, it may give learners the feeling of being watched by the speaker, which, in turn, may increase their level of anxiety (cf. Chapter 10.4). As a result, the profile kept by the virtual native speaker involves maintaining an appropriate level of presence in the virtual room. The speaker has to be available without distracting or disturbing the learner. When there is nothing to do for the virtual native speaker, it may choose to occupy itself otherwise in the virtual room; however, it should stay in the room to meet the requirement of constant availability. Whatever the virtual native speaker does when it is idle should not draw the learner's attention too much in order to avoid causing distraction. Furthermore, the speaker should interrupt its current idle-state occupation when the learner wants to interact with the agent again (it should be noted that this is a higher level of cooperativeness than the learner could expect from a human informant). Finally, even while serving its main function as the learner's informant, the virtual native speaker should not draw more attention to itself than is appropriate for its role, especially when other agents (other virtual native speakers, the coach, or peer agents) are present.

16.3.2.4 Relationship-Building Behaviors

The virtual native speaker interacts with the learner both regularly and over longer periods of time in the course of a fieldwork class. This provides opportunities for the speaker to work on its relationship with the learner in order to develop it into a productive, cooperative partnership.

One obstacle to relationship building is the fact that the learner does not know the language of the virtual native speaker (yet), and vice versa. Therefore, verbal relationship-building strategies, such as engaging the learner in small talk (cf. Chapter 10.3), are not a very promising option for the speaker.[265]

However, there are other ways for the virtual native speaker to reach out to the learner. As a first step, before any regular sessions between the speaker and the learner take place, they meet and become acquainted. In this first meeting, it is critical for the virtual native speaker to signal a positive attitude toward the learner through its face, body, and voice, in order to contribute to a favorable first impression. Some other individual (e.g. the coach agent as the master teacher

[265] The virtual native speaker could resort to an auxiliary natural language, such as English, for its social interactions with the learner. In principle, there is no reason why the speaker should know only its native language. Still, it is already difficult enough to develop language processing capabilities for a single language without adding a second (although processing components for a common language like English may be readily available). Furthermore, the same voice would have to be available for both the first and the second language. The English voice would also require an appropriate foreign accent. All this is hard to realize with speech synthesis technology. On the positive side, since the auxiliary language is not the virtual native speaker's first language, the agent might be able to get away with shortcomings of its linguistic performance, which are unavoidable given the current state of human language technologies (cf. Chapter 11.1).

of the apprentice learner, cf. Section 16.2.3) may introduce the learner and the virtual native speaker to each other, or they introduce themselves. For example, this can happen during a virtual warm-up meeting, where the learner meets the members of his or her hybrid fieldwork community, including the human class instructor, the other human class participants, the learner's personal coach agent, and the peer agents (cf. Section 16.2.4) that will take the fieldwork class together with the learner. Apart from names, the learner and the virtual native speaker may exchange further information about themselves for the purpose of getting to know each other. One way for the virtual native speaker to obtain some personal information about the learner is to access the learner's user profile on the Virtual Linguistics Campus. The advantage of this method is that the learner can decide beforehand which information about himself or herself is visible to other individuals, including the virtual native speaker. Another possibility is for the learner to answer a small set of closed questions about his or her personal background, his or her reasons for taking the fieldwork class, his or her expectations, and so on. In general, if the learner chooses to share personal details with the virtual native speaker, whether at the beginning or later in a fieldwork class, he or she must be able to trust the speaker to keep this information confidential. In case the learner does not wish to disclose information about himself or herself, the virtual native speaker should accept that decision and treat the learner with the same respect as otherwise.

Whether or not the learner wants to share personal information with the virtual native speaker, for the latter to play its part in the relationship-building process and to gain depth and credibility as an individual, it is vital that a rich personal back-story of the virtual native speaker is (gradually) revealed to the learner (see Section 16.3.5.3 for a detailed discussion). It should be emphasized at this point that the virtual native speaker is not a copy of the human native speaker who provided the voice talent. Instead, it has a personality and a (virtual) life of its own, which progresses and changes in the course of the fieldwork class (cf. Section 16.3.5.2 and Section 16.3.5.3).

In the fieldwork sessions following their introduction to each other, it is important for the growing relationship between the virtual native speaker and the learner that the speaker remembers the learner and its previous interactions with him or her (including exchanges of personal information). When the learner enters the virtual room of a fieldwork unit for the first time on a given day, the virtual native speaker should acknowledge the learner's presence by turning toward him or her and welcoming the learner in its native language. On return visits of the learner, the greeting should be adapted accordingly. While the speaker's way of greeting the learner may be more formal and reserved in the early stages of their relationship, it can change to less formal and more intimate forms over time.

If the speaker is occupied with something when the learner enters, it should stop this activity as soon as possible and tend to its visitor (cf. Section 16.3.2.3). In fact, the virtual native speaker could treat the learner as a guest and show him or her hospitality according to the customs of its culture (although the options for offering things such as food and drink to the learner are obviously limited). In the beginning, when the relationship between the speaker and the learner is not that close yet, the speaker's behaviors may still indicate some reserve, which may be expressed through its voice, facial expressions, gaze patterns, and posture (modulated by the rules for expressing interpersonal relationships in the culture represented by the virtual native speaker). The reserved speaker's voice may be polite but not too warm; its smiles may be less frequent and intense; it may maintain less eye contact; its stance may be closed and relaxed; and it may stay further away and lean away from the learner (cf. Table 14). During the following sessions, as familiarity and liking (at least on the part of the virtual native speaker) increase, these behaviors gradually change toward friendliness to reflect the closer relationship with the learner (Bickmore et al. 2005). A virtual native speaker that is familiar with and likes the learner may have a warm, energetic voice and a smiling face, maintain steady eye contact, show an open and relaxed stance, and close the distance between itself and the learner (again subject to cultural modulation). Unfortunately, the virtual native speaker is unable to determine if the learner reciprocates because the technical possibilities for obtaining feedback, in particular non-verbal feedback, from the learner in a web-based environment are restricted (cf. Chapter 7.10). However, given the general hedonic preference of people for positive expressions (of emotion) shown by others (cf. Chapter 13.4.2), the displays of increasing friendliness by the virtual native speaker may be well-received by learners.

16.3.3 Conversation

The current virtual native speaker produces spoken utterances in response to the learner's point-and-click actions in the virtual room. In contrast, the enhanced speaker agent is a conversational agent (cf. Chapter 11.3) that is equipped with the capacity to participate in a wider range of pre-scripted and dynamic exchanges with other virtual actors and with the learner as well as with the ability to produce more varied utterances. The verbal messages of the virtual native speaker (and other virtual actors) are integrated with its other behaviors for communication, enhancing believability, and relationship building, which were described in the previous sections. The interactions and performances taking place in the virtual room are recorded by the virtual fieldwork environment and can be played back by the learner as often as necessary. The learner can also modify or script performances, including dialogues (cf. Section 16.3.4).

The function of the virtual native speaker's enhanced communicative abilities is first and foremost to expose the learner to more and more varied linguistic data from the language of the fieldwork class. The domain of the conversations in this language is limited to the environment of the virtual room. The interlocutors (which may include virtual native speakers as well as the learner) talk about the objects and events in the virtual room which they observe or in which they participate. The complexity of the utterances occurring in these dialogues is restricted to a comparatively simple level (e.g. a single declarative or interrogative sentence) that is within or just beyond the learner's present knowledge and may expose him or her to a new phenomenon of the target language that is the focus of the current fieldwork unit (e.g. word-order or question formation) while keeping the amount of new linguistic material to be collected and analyzed by the learner from the utterance of a speaker to a manageable size. In the early units of a fieldwork class, the exchanges may be as simple as a word or phrase produced in response to a question like "What is X?" (to make the virtual native speaker identify an object) or "Where is Y?" (to elicit utterances illustrating how spatial relations are expressed or whether the language uses prepositions or postpositions). In order to make its responses to such queries unambiguous for the observing learner, a virtual actor uses believable deictic behaviors (cf. Section 16.3.2.1.2), which may involve pointing gestures, locomotion, eye gaze, and other non-verbal behaviors to support its verbal utterance.

Learners need facilities for formulating queries to the virtual native speaker or to a team of speakers (and for otherwise participating in interactions taking place in the virtual room). The most sophisticated option allows the learner to use combined speech and mouse input within a multimodal interface (cf. Chapter 11.2) to the virtual room. In the beginning, when the learner is new to the language, he or she will make more use of the mouse to build queries. In the simplest case, the learner may select an object with the mouse and then pick a type of query from a menu, such as "What is X?," "Where is Y?," and "What color is Z?." If there are several actors present that can be queried, the learner can simply click on the addressee of his or her question. In the following, as the learner is picking up more and more vocabulary and constructions along the way of exploring the target language, he or she may feel increasingly confident (and may also be encouraged) to speak the language in order to ask questions and make other contributions to conversations in the virtual room. Mouse-based gestures may still be used to clarify verbal meanings, such as the referents of objects. As a useful side effect, the learner's utterances can be assessed with respect to their grammaticality and accuracy of pronunciation,[266] which adds further aspects of language courses to the fieldwork classes (although their focus remains on linguistic fieldwork rather than language learning).

However, equipping the virtual native speaker with the ability to interpret multimodal speech and mouse input is ambitious implementation-wise. While the interpretation of mouse-based gestures is quite straightforward in the context of the virtual room, especially if supported by appropriate menu options (see below), speech recognition (cf. Chapter 11.1.1) and natural language understanding (cf. Chapter 11.1.3) still pose considerable challenges with respect to working around their present limitations and integrating them into a web-based environment (cf. Chapter 14.3.2).

Similar to speech synthesis systems (cf. Section 16.3.2.1.1), speech recognizers are available for English, German, and other more common languages but not necessarily for the less common ones, such as Welsh (Jones 2003). In addition, speech recognition products are often trained to recognize the speech of proficient speakers of the language and may require considerable adjustments to enable recognition and diagnosis of the less than perfect spoken inputs from fieldwork learners. On the other hand, both the vocabulary size and the perplexity (cf. Chapter 11.1.1.1) of the speech recognition task in the virtual fieldwork environment are way below those of a general-purpose dictation system. However, the speech recognizer would have to be speaker-independent and should permit continuous input. Concerning the technical integration of speech recognition into the web-based fieldwork environment, there are three options: as a client-side plug-in for the web browser, as a separate speech recognition server, or as a hybrid of these two approaches, which divides the workload between the client side and the server side. The first option has the advantage of reducing the network traffic and the workload on the server side, but it is dependent on the capabilities of the learner's computer and the successful installation of the necessary components on that machine. The second option minimizes local installation and performance problems, but the network between the speech recognition server and the learner's computer may become the performance bottleneck, leading to perceptible latencies that disrupt the immediacy and continuity of responses from the virtual native speaker and annoy the learner. The third option addresses the issues of installation, performance, and network latencies by trying to find an appropriate distribution of processing components between server and clients (cf. Chapter 14.3.2.1). This hybrid strategy may also be appropriate for the other components of the virtual native speaker as an embodied interactive software agent (cf. Chapter 3.3).

As discussed in Chapter 11.1.3.5, the ability of machines to interpret natural language utterances above the level of individual sentences is still quite limited. With current NLU technologies, single sentences can be analyzed morphologically, syntactically, and semantically to produce a logical form or a conceptual structure of the sentence, whereas the interpretation of

[266] This assessment is not performed by the virtual native speaker because foreign-language tutoring is outside of its domain of expertise. It could be the responsibility of a separate language-tutor agent or of some kind of disembodied intelligent tutoring component, as in the Tactical Language and Culture Training System (cf. Footnote 261).

the sentences with respect to their role in the discourse and the communicative intentions of the speaker or writer is much more difficult. In general, the deeper the desired level of analysis, the more it may be necessary for developers to narrow the domain. In the context of the virtual room, the use of natural language understanding may be feasible, though, because the domain is quite small (being essentially limited to the actors, objects, and events of the virtual room) and the utterances that have to be analyzed are comparatively simple (usually consisting of single sentences that follow a limited set of patterns and involve a small vocabulary). However, the NLU process must be prepared to handle ill-formed input and to diagnose the mistakes in the learner's utterances (so another tutoring agent or system can address them). Concerning the implementation of the NLU functionality, the same problem arises as with the other human language technologies involved in the design of the virtual native speaker: it may be difficult to obtain the required knowledge sources (grammars, lexicons, etc.) and processing components (morphological analyzers, syntactic parsers, etc.) for the language of a particular fieldwork class because no one has considered building them.

Apart from developing the processing of spoken language and mouse input as individual capabilities, it is also necessary to integrate and coordinate these two modes when the learner uses them in combination. Since speech and mouse actions are less coupled temporally than, say, speech and lip movements, and provide complementary information, the two input modes can be processed in some depth in parallel first and then integrated (fused) later at a higher semantic level (cf. Chapter 11.2).

At present, multimodal input, while technologically possible, is difficult to implement, in particular as regards the speech and language processing components and their integration into the environment and with mouse input. Hence, it is more feasible to consider combined speech and mouse input as an option for future versions of the fieldwork classes and to rely for the time being only on mouse input. Given that spoken language input is not crucial for the learner's interaction with the virtual native speaker (after all, the speaker produces utterances for the learner, not vice versa), this is an acceptable compromise, although it means an asymmetry between the input modes available to the learner and the output modes available to the virtual native speaker (cf. Chapter 7.10): while the speaker uses spoken utterances in the interaction, the learner does not have this option.

In order to increase the flexibility of the learner to construct all kinds of queries with the mouse, to which the speaker responds with dynamically generated utterances describing the learner's constructions (see below), a wider range of mouse-based actions should be possible in the virtual room, including:

- Selecting both individual and multiple objects and actors;
- Grouping objects and actors;
- Changing the properties of objects, actors, and events;
- Dragging objects and actors to different locations;
- Combining objects, actors, and activities into events;
- Specifying the locations of events;
- Sequencing multiple events.

The larger inventory of mouse actions is supported by an extended set of menu options that allow the learner to specify what types of actions or utterances should be performed by the virtual actors given the objects, actors, and events that have been arranged with the mouse. Examples include:

- Label the selected object or actor.
- Describe the specified event or arrangement of objects and/or actors.
- Perform the specified event or sequence of events.
- Ask the selected actor about a particular aspect of the current performance.
- Discuss a particular event with other actors.
- Refer to previous and future events and states of the virtual room.

The resulting user interface to the virtual room provides many more options for learners than the current interface. Therefore, it is important that the enhanced functionality does not increase the complexity of using the interface (cf. Chapter 6.3). Ideally, the learner should not require instructions on how to interact with the room and the virtual actors; if they are necessary, they could be provided by the learner's personal coach agent (cf. Section 16.2.3) as part of introducing the learner to the virtual fieldwork environment.

The ability of the virtual native speaker to respond flexibly to the learner's actions and inquiries in the virtual room requires a natural language generation component (cf. Chapter 11.1.4) for the language of the fieldwork class, which dynamically generates the appropriate utterance for the current context from an internal non-linguistic representation derived by processing input from the learner or other virtual actors. Given that natural language generation systems are still not widely available, even for the more 'important' languages, it is very likely that this component will have to be a proprietary development. However, it is not necessary to create a full-fledged natural language generator with the ability to produce unrestricted linguistic outputs in the target language. The length and complexity of the utterances that have to be generated are restricted by the goal and environment of the fieldwork classes. Typically, the utterances of a virtual native speaker will consist of a single sentence with a structure of limited complexity and a small vocabulary. The generative capabilities of the speaker have to be sufficient to create the set of utterances that describe the range of objects, actors, arrangements, and events which are (or were or will be) present in or can be created in the virtual room. In general, the utterances of the virtual native speaker should cover the range of linguistic aspects to be explored by the learner in the fieldwork class in adequate breadth and depth, providing enough raw material to cover all phenomena of interest and each individual phenomenon in sufficient detail.

While natural language generation components typically produce text output, the utterances of the virtual native speaker have to be delivered in spoken rather than written format. Text output in its native language is not an option for the virtual native speaker because learners need the spoken utterances to uncover the sound system of the language, given that the relationship between spelling and pronunciation is quite unpredictable in human languages. In addition, the language of the fieldwork class may not have a writing system, or it may have one that uses characters which are unknown to the learner. Finally, it is the spoken rather than the written form of language that is regarded as primary by modern linguistic science. For these reasons, the natural language generator has to be coupled with a speech synthesis system that delivers the utterances of the virtual native speaker in spoken format (cf. Section 16.3.2.1.1).

Unlike the dialogues between other types of conversational agents and human users, the interactions between the learner and the virtual native speaker are open-ended, in the sense of not being directed toward achieving a particular pre-set goal, such as booking a flight or solving a computer problem. Instead, the continuing goal of the exchanges between the learner and the speaker is to elicit more and more varied linguistic data from the virtual native speaker in the context of the topic of the current fieldwork unit. The interactions do not develop along a predictable path because the speaker does not know what situation the learner might refer to or create next in the virtual room. But the virtual native speaker nonetheless has to keep track of its interactions with the learner and with other virtual actors, both during the current fieldwork session and across the series of sessions in the current fieldwork class, in order to know what has been talked about before and what entities (objects, actors, and events) are currently in focus, so it can generate the appropriate utterances. Furthermore, it is important for the speaker to maintain a history of previous interactions and performances because the learner should be able to replay them in order to hear the utterances illustrating the relevant linguistic phenomena again in the context in which they originally occurred.

In the current implementation of the virtual native speaker, the learner has the initiative in his or her interactions with the speaker while the latter patiently waits for him or her to request the next utterance. The fieldworker-informant relationship largely defines the pattern for the interactions between the learner and the enhanced virtual native speaker as well, including the allocation of control. Since the purpose of these exchanges is to give the learner the opportunity to elicit data from the virtual native speaker (although the roles might be reversed, cf. Section 16.2.2), the learner will be in control of the interaction most of the time. However, the virtual native speaker might also take the initiative occasionally and make contributions proactively (cf. Chapter 3.1.6), including both utterances and performances, that provide further data for the learner's analysis of a given phenomenon or even expose him or her to a new aspect of the target language, giving the learner's exploration of the virtual room a new direction (and hence some implicit guidance).

While learner initiative (or user initiative, cf. Chapter 11.3.3.2.1) characterizes the exchanges between the learner and the virtual native speaker, the virtual actors involved in dialogues as part of their performances in the virtual room share control over their interactions, each of them being able to take the initiative as appropriate. The structure of these dialogues may be given by the script or plan for the performance (cf. Section 16.3.4) or may develop as the interaction progresses (when the learner is involved). In both cases, the actual utterances occurring in the dialogue are dynamically generated by the virtual actors based on the script/plan or the current state and requirements of the interaction. To the observer, the exchanges between the virtual actors may give the appearance of an actual dialogue; however, it is not necessary for a virtual actor to use natural language understanding technology in order to interpret the contributions of other actors. Either the meaning of the utterances is given in the script or plan for the performance, or the actors internally exchange messages in an agent communication language (cf. Chapter 14.1.5) while producing the appropriate natural language utterances for the current stage of the interaction.

16.3.4 Expertise

To be useful for learners, pedagogical software agents should possess substantial and relevant expertise that they can offer to learners (cf. Chapter 7.1 and Chapter 12). The virtual native speaker provides essential expertise in its native language and culture for learners exploring the virtual room of a fieldwork class to uncover the structure of this language and the conventions and customs of the culture in which it is spoken. Without the virtual native speaker, learners would not obtain linguistic data for their analysis. Furthermore, their opportunities for learning about the culture would be severely reduced. The virtual native speaker does not have to be able to teach its native language to learners or coach them in the linguistic fieldwork methods necessary for its analysis because these tasks are outside the scope of its role (although the speaker's utterances can highlight particular features of the language and may even guide the learner toward new ones, cf. Section 16.2.2). Instead, the expertise of the virtual native speaker manifests itself primarily in the ability to produce utterances in its native language which are authentic as regards pronunciation, vocabulary, and word and sentence structure, and which accurately and appropriately describe states of affairs in the virtual room.

The learner relies on the authenticity and accuracy of the virtual native speaker's utterances, so building this part of its expertise requires special attention and care. Components, resources, and models for speech and language processing in combination with non-verbal communicative and other behaviors have to be developed, which allow the speaker agent to put all the events and states that are possible in the virtual room into spoken utterances and to accompany them with contextually appropriate gestures, facial expressions, gaze, posture, and locomotion. Since the emphasis of the speaker's role is on language production rather than understanding, and the option for learners to address the speaker or other virtual actors with spoken inputs in the language of the fieldwork class is desirable but not crucial, development may focus, at least for the short term, on the natural language generation and speech synthesis capabilities of the virtual native speaker and on their coordination and integration with the speaker's non-linguistic behaviors. The design of these aspects was discussed in depth in the previous section.

In addition, the virtual native speaker requires detailed knowledge about its environment to make sure that its utterances (and other behaviors) correctly reflect the current (or some previous or future) state of affairs in the room. Given that the enhanced virtual room will be a dynamic environment offering the learner, the virtual native speaker, and other virtual actors many more options to manipulate the objects, actors, and events in the room than before, any attempt to anticipate and pre-define the entire range of potential situations in the virtual room is unlikely to succeed. Instead, it is necessary to maintain a dynamically updated model of the entities and relationships in the virtual native speaker's environment (including the learner and his or her actions), from which relevant information is selected and organized in order to generate utterances that describe the situation of interest. Furthermore, as competence in a language also involves the ability to talk about the past, the future, and potential or hypothetical actions and situations, the speaker's model should extend beyond the current state of the virtual room and include knowledge about what was, what will be, and what might be taking place in the room. While maintaining this model would be a formidable task for large and complex environments, the relatively small setting of the virtual room makes it possible to manage the complexity of the interactions and of the modeling required (although this is likely to change if the virtual room is extended to a virtual world, as proposed in Section 16.2.2).

The model of the virtual room could be implemented as an internal representation that is built only for the virtual native speaker and cannot be accessed by others, but it seems more appropriate to externalize this representation and encode it as part of the environment (the virtual room), using the languages of the Semantic Web, as proposed in Chapter 12.1.2.5. In this approach, semantic markup is added to the virtual room, which annotates the room, its entities (objects, actors, and events), and their relationships in terms of metadata statements whose terminology is provided by a shared ontology created for the linguistic fieldwork classes. New entities added to the virtual room are described by their own semantic metadata which integrates seamlessly with the existing semantic markup because it uses the same formalism and the same ontology. Changes to an entity in the virtual room modify its markup. A history of changes is maintained for the entity, which allows the virtual native speaker to refer to a previous location of an object, for example.

The use of an external model of the virtual room based on semantic metadata has a number of advantages over an internal representation. First, it is not necessary to maintain separate models for different virtual actors. The individual agents only require the ability to read and process the semantic annotations associated with entities in the virtual room. They do not have to encode and keep track of their properties or status themselves. Second, since the virtual room has the same layout and inventory in all fieldwork classes, it becomes possible to create a single shared ontology that describes the structure of this domain. This ontology will be quite small because the virtual room only contains entities of a limited number of types. Third, the whole metadata and ontology mechanism is based on logic (cf. Chapter 12.1.2.1), which allows agents to process retrieved metadata and deduce new facts from it through the application of inference rules. The new facts can then be used by the virtual native speaker to talk about aspects of the virtual room in its utterances that are not explicitly encoded in the metadata structures.

The metadata statements describing the virtual room have to be organized and stored in a way that facilitates their controlled retrieval, modification, addition, and removal by different (distributed) parties (learners, developers, and virtual actors), both within and across fieldwork sessions. Basically, the metadata model of the virtual room could be maintained on the server, on the learner's computer, or both. If each learner works with his or her own view of the virtual room independently of others (i.e. the virtual room is not a shared environment), a local copy can be maintained during the fieldwork session and stored on the server after the session. However, if the virtual room is shared by several learners at different computers, the issue of synchronizing the different local copies of the model of the virtual room needs to be addressed.

Apart from providing learners with authentic and accurate linguistic data, a further aspect of the virtual native speaker's expertise concerns its ability to participate in virtual performances that may involve other virtual actors, objects in the virtual room, and possibly the learner (cf. Chapter 12.3.12). One way to create these performances is to author them in advance and have the virtual actor(s) perform them later. Performances can be scripted by learners, instructors, and developers, using a *visual authoring tool* (to be developed) that provides an intuitive graphical interface for modeling the scene to be performed. This tool generates a script for the performance that specifies the actors, roles, props, location, and the temporally ordered sequence of actions to be carried out (cf. Chapter 14.2). The script can be stored in the system to be executed later, either at the learner's request or proactively initiated by the virtual native speaker or by the coach agent. At run-time, the character player executes the script to create (sequence) the animations of the virtual actors plus the necessary animations of objects in the virtual room, e.g. the opening of the door or the ball bouncing off the wall, cf. Figure 94). The visual authoring tool should give learners and instructors the possibility to author character performances at a high, non-technical level and leave the details of coordinating and sequencing the verbal and non-verbal behaviors to the behavior generation component. Developers may be given more control over the scene by allowing them to define more fine-grained details of the characters and the performance.

It is also possible for learners to initiate dynamically generated virtual performances by specifying a particular linguistic concept from the target language (e.g. SVO-sentences[267] or wh-questions[268]) plus a number of parameters (actors, props, etc.) and having the virtual actors improvise a scene that illustrates the concept through the utterances and behaviors produced by the actors, based on a dynamically constructed plan for the performance and the embedded

[267] SVO is an abbreviation for "Subject-Verb-Object" and labels a particular linear ordering pattern of constituents with these syntactic functions in sentences.

[268] *Wh-questions* are interrogative sentences whose first constituent (the *wh-element*) begins with or consists of one of the following interrogative words (called *wh-words*): *who*, *whom*, *whose*, *what*, *which*, *when*, *where*, *how*, and *why*). They are typically posed to elicit information of the kind indicated by the wh-word from the addressee.

dialogue (cf. Chapter 11.3.3.2.2 and Chapter 12.1.3.3), whose building blocks are provided by a library of plans, partial plans, and plan operators. Furthermore, an actor or a group of actors may take the initiative, develop a plan for illustrating a phenomenon of the language or a convention of the culture, and stage their own performance by executing this plan, without directions from the learner or other parties.

An important feature of the virtual performances is their interactivity: learners can interrupt a performance to query the actors about aspects of the scene, modify elements of the performance and observe the results, and even participate in the performance via (combined) speech and mouse input (cf. Chapter 16.3.3). If the learner interrupts a performance in progress with a query, the execution of the underlying script or plan should stop at the next appropriate point (which means that the current utterance or other action may have to be finished first). After dealing with the learner's query, the performance should continue where it left off, provided that the learner does not choose to abort it. The situation becomes more complicated when an ongoing performance is modified by the learner. The first concern is to check if the modification is legal, i.e. does not violate any constraints or lead to undefined states of the environment. If a legal modification happens while the performance is halted, the plan for the remainder of the performance has to be updated accordingly before the actors continue. If the learner changes something during the performance, the necessary modifications to the plan have to be made on the fly. Depending on the nature of the learner's manipulation, minor or major plan revisions may become necessary, even to the extent that the current plan becomes obsolete and a new one has to be generated. In the latter case, the performance cannot continue, and the actors have to start again. Other revisions may be less radical, but it may still be necessary for the actors to go back and redo earlier parts of the performance. To simplify the task of the planner, the range of possible modifications to a performance could be restricted to those that do not render aspects of the plan inappropriate that have already been executed. Handling modifications to script-based performances presents another challenge because unlike plans, scripts are static representations and therefore cannot be changed on the fly. A simple solution would be to modify and re-execute the script. However, repeating the performance every time a change is made can be tiresome for the learner. Alternatively, upon execution, the script could be converted into an internal plan which, in turn, could be flexibly changed at run-time.

The most challenging scenario for the virtual performances arises when the learner is given the opportunity to become involved in a performance as an actor playing his or her own part (rather than as a director or inquirer), in which he or she interacts both with virtual actors and objects, using mouse-based actions and/or speech input (in the future). In these performances, not just the interactions among the virtual actors but also between the virtual actors and the learner have to be modeled (cf. Chapter 12.3.12). In the simplest case, the script or plan for the performance specifies the learner's part completely, including his or her options for utterances and other actions. At each stage of the performance, the learner selects (or speaks) an utterance or an action from the limited set of options made available. However, this restricted form of participation does not give the learner the freedom to produce his or her own utterances and to influence the direction of the performance, reducing his or her opportunities for practicing and experimenting with structures of the language. If the restrictions on the learner's possibilities for participation are removed, virtual performances and their embedded dialogues may develop unpredictably (although within the boundaries of the virtual room). As a result, there can be no advance script or plan for a performance; rather, the plan evolves as the human and virtual actors move on with their interaction. The evolving plan integrates the goals and plans of the virtual actors with those inferred from the learner's observed actions and utterances during the performance.

To make these hybrid interactive performances manageable, several restrictions can be applied. First, the virtual room provides visual clues (objects, actors, and events) that indicate potential actors, roles, props, and plots. Furthermore, the virtual actors can perform utterances and behaviors that strongly suggest how the learner should respond, guiding the performance toward topics and patterns that can be handled. Finally, the limited knowledge that the learner has been able to pick up about the language from the previous fieldwork sessions restricts the range of his or her contributions to a performance. In fact, everything the learner knows about the language comes from the virtual room, the very domain of the performance in which he or she is participating.

Allowing for human authoring of interactions in the virtual room involves the risk that the resulting performances become too long and lack focus. However, as a general requirement, whether pre-scripted or dynamic, each performance delivered in the virtual room has to be short and should focus on a small set of linguistic concepts of the target language that are determined by the topic of the current fieldwork unit. In other words, it is necessary to build restrictions with respect to focus and duration into both the authoring tool and the performance planning component. The performance-authoring feature should be used as a research tool rather than as a toy. After all, the purpose of the virtual performances is not to entertain the learner or even distract him or her from the fieldwork tasks but to expose him or her to further data from the target language. This is not to say, though, that the virtual performances, including the dialogues, should be boring and repetitive (except when a replay of a performance has been requested by the learner). In fact, each of the actors involved may choose to add its own touch to a performance, which includes a distinctive verbal and non-verbal style.

The virtual native speaker has regular and extended interactions with a particular learner in the course of a fieldwork class, during which the learner elicits data from the speaker. To increase the effectiveness of these interactions and to promote its relationship with the learner across a series of fieldwork sessions, the speaker should be able to adapt its behaviors to the individual learner over time. In order to adapt to the learner, the virtual native speaker collects information about the learner and uses it to construct a long-term student model (cf. Chapter 12.2). The modeling starts from a set of assumptions about a typical learner of the linguistic fieldwork classes on the Virtual Linguistics Campus and successively refines these assumptions as the speaker becomes more and more acquainted with the individual learner and his or her preferences and habits. As the fieldwork class goes on, the following information gradually accumulates in the student model:

- Personal details of the learner (e.g. name, sex, country, reasons for taking the fieldwork class, expectations regarding the class and the virtual native speaker, etc.);
- The learner's preferred input mode (mouse, speech, or a combination of the two);
- Adaptations of the speaker's behavior (e.g. speech rate, speech volume, animation speed, repetition of utterances (at a slower rate) by default, utterance length and complexity, etc.);
- Typical patterns underlying the learner's interaction with the virtual native speaker, other virtual actors, and with the virtual room;
- Inferred recurring strategies of the learner for exploring the virtual room;
- Types of situations, events, and performances created or manipulated by the learner (e.g. single-actor vs. two-actor vs. multi-actor performances);

- Linguistic data (and the underlying concepts of the language) to which the learner has already been exposed, both in the current and in previous sessions, through the speaker's own utterances and those provided by other virtual actors;
- Cognitive and emotional states of the learner (e.g. boredom, excitement, frustration, etc.).

Unlike the student models found in intelligent tutoring systems or pedagogical tutor agents (cf. Chapter 12.2), the virtual native speaker does not have to model the learner's developing knowledge of the language of the fieldwork class or his or her skill level in linguistic fieldwork, including any misconceptions that the learner may have in either area (modeling the learner's knowledge and skills is of greater value for the coach agent, cf. Section 16.2.3), although it could borrow this information from the coach agent to be able to present more challenging linguistic data to advanced learners and to simplify its utterances for novices or learners with low ability.

Some of the information for building the student model is specified by the learner (e.g. his or her preferences regarding input mode and speaker behavior) while other aspects can be obtained from external sources (e.g. the learner's VLC user profile or the student models built by other agents). Still others have to be inferred from observations of the learner's behavior (e.g. the learner's internal state, interaction patterns, and exploratory strategies), which can be difficult because only limited information about the learner can be reliably collected from the few input channels that are available in a web-based environment (cf. Section 16.3.5.1.2).

16.3.5 Emotion and Personality

With the ingredients described in the previous sections, the virtual native speaker becomes a competent and communicative actor in the virtual room that can interact verbally and non-verbally with the learner and with other virtual actors as well as participate in virtual character performances in the room. However, the speaker's "inner life" still requires some work for the agent to be perceived as an emotionally adept, appealing, and deep individual. Three aspects are discussed in the subsections below: the virtual native speaker's ability to model, interpret, and express emotions, including its own as well as those of the learner and other virtual actors; the design of the speaker's personality; and the crafting of its individual back-story.

16.3.5.1 Emotion

The role of the virtual native speaker requires an embodied pedagogical software agent to serve the learner as an effective informant and performer, to project a believable agent persona, and to build a productive relationship with the learner. Each of these three aspects (task, persona, and relationship building) involves the need for the agent to handle emotions (cf. Chapter 13.4) in a competent way, given its own restrictions and those of its environment. Emotional intelligence is important for the virtual native speaker's task because it makes interactions and performances more lively, interesting, and informative. It increases the believability of the speaker's persona because an agent that expresses (and responds to) emotions is more likely to be perceived as a sentient and sensitive individual rather than as a cold and uncaring automaton. Finally, it facilitates relationship building because the learner perceives that the virtual native speaker cares about him or her. Emotional competence includes the three capabilities of emotion modeling (cf. Chapter 13.3), emotion interpretation (cf. Chapter 13.4.1), and emotion expression (cf. Chapter 13.4.2), which are discussed in the following.

16.3.5.1.1 Emotion Modeling

The purpose of a computational model of emotion is to provide a symbolic representation of emotion and emotional dynamics that facilitates the recognition, interpretation, disambiguation, management, and synthesis of emotions in an agent's application context (cf. Chapter 13.3). In principle, the virtual native speaker can model the emotions of the learner and the other virtual actors as well as its own emotional state and use this information to understand and predict the emotional states and reactions of other individuals in the fieldwork environment and to come up with its own emotional responses to them and to its own appraisal (cf. Chapter 13.1.2.1) of situations and events in the virtual room.

Computational models of emotion differ with respect to whether they focus on enabling the deliberate communication of emotion to achieve certain communicative or manipulative effects (cf. Chapter 13.3.1), or attempt to give an agent 'true' emotions by simulating aspects of emotion processes (cf. Chapter 13.3.2). For the virtual native speaker, a communication-driven approach to emotion modeling is adopted because it is more important for the speaker to be able to successfully communicate contextually, culturally, and individually appropriate emotions to the learner than to simulate internal emotional mechanisms and states. The more sophisticated simulation-based emotion models discussed in Chapter 13.3.2 would allow the virtual native speaker to appraise (evaluate) states of affairs in its environment more deeply and accurately and offer different (problem- and emotion-focused) strategies for coping with them. However, given the comparative simplicity of the virtual room and the interactions taking place in it, as well as the pragmatic emphasis on the social utility of the speaker's surface emotional behavior (for communication, enhancing believability, and relationship building), rather than the adaptive functions of complex internal mechanisms, the complexity of simulation-based models outweighs the appeal of endowing the virtual native speaker with genuine (simulated) emotions.

16.3.5.1.2 Emotion Interpretation

Effective social functioning of agents requires emotional intelligence (cf. Chapter 13.4), which includes the ability to interpret the emotions of the user, other agents, and the agent itself. For the virtual native speaker, emotional intelligence is useful to understand the behavior of the learner and other virtual actors and to respond with appropriate emotional expressions of its own (cf. Section 16.3.5.1.3). Unfortunately, while the speaker has access to the emotional states of other virtual actors, its possibilities to determine the learner's state are limited, due to the restricted range of input modes that learners can use (or choose to make use of) in a web-based environment (cf. Chapter 7.10), the continuous and variable nature of the auditory and visual input signals, and the shortcomings of the recognition technologies required. The most feasible input option is still the mouse (and the keyboard, but the latter is not used very much in the virtual room). However, when only the learner's mouse actions are available for analysis, it becomes very difficult to deduce the learner's emotional state from his or her action patterns (e.g. click paths and click frequencies). For example, does the learner's repeated request for the same utterance mean that he or she does not know how to analyze it and begins to feel frustrated? Does erratic clicking indicate anger or confusion? How should longer periods of inactivity be interpreted? Is the learner thinking, unsure about how to proceed, or did he or she leave? Without a camera and a microphone, the virtual native speaker is essentially 'blind' and 'deaf' as far as the learner's actions outside the virtual room are concerned. In future versions of the virtual fieldwork environment, the auditory and visual channels may be available as additional input options to broaden the range of emotional cues that can be collected from the learner's face and voice (cf. Table 23 and Table 26). These cues will then have to be integrated, both with each other and with the learner's mouse input (cf. Chapter 11.2). Some pedagogical agent projects already experiment with the use of multiple input channels for emotion interpretation. For example, "AutoTutor [(cf. Figure 17d)] has been enhanced to identify and assess learners' affective states through dialogue assessment techniques, video capture of the face, learner body posture, and a pressure sensitive mouse and keyboard" (Graesser et al. 2007). While using advanced input devices is possible in experimental setups in the laboratory, it is not an option for a real-world e-learning environment, where developers have to consider the devices that are available to all or at least the majority of learners.

Overall, for the time being, the insecurity of the virtual native speaker, and the coach and peer agents as well, with respect to how they should behave toward the learner if they cannot reliably detect the learner's emotional state, remains unresolved. Having said that, the direct approach of explicitly asking the learner about his or her feelings is always open to all agents in the fieldwork environment.

16.3.5.1.3 Emotion Expression

While the virtual native speaker's options for emotion interpretation are limited, on the output side, it has the whole spectrum of verbal and non-verbal behaviors enabled by its face, body, and voice at its disposal to create emotional displays for the learner that combine visual and vocal signals. As a general requirement, the emotional expressions of the speaker (or any other agent in the fieldwork environment) have to be both easy to interpret and believable but at the same time should also be culturally and contextually appropriate.

The primary devices for conveying emotions are the face and the voice, but gaze, gesture, posture, and locomotion also provide visible cues as to the speaker's emotional state (cf. Chapter 10.1 and Chapter 13.4.2). The facial expressions of the virtual native speaker can be described in terms of different shapes of the eyes, eyebrows, and mouth (cf. Section 16.3.2.1.3). A more elaborate system for the classification and specification of human facial expressions is provided by the Facial Action Coding System (cf. Table 24 and Table 25), but the resulting descriptions would possibly contain more detail than necessary for controlling the emotional expressions of the cartoon-like face of the speaker character.

Similar to the face, different aspects of the speaker's voice provide emotional cues, including speech rate, pitch average, pitch range, intensity, voice quality, pitch changes, and articulation (cf. Table 26). However, while animated faces can believably portray a variety of emotions, the vocal expression of emotion using synthetic speech is more difficult (just like the generation of natural language utterances with emotional content), although the expressive capabilities of machine-generated voices are developing and people have been demonstrated to be able to detect emotions in synthetic speech with an accuracy that is not too far below the recognition rates for emotions in human voices (cf. Chapter 13.4.2).

Neither the face nor the voice can convey the full range of emotions unambiguously on its own. In fact, both facial and vocal expressions of different emotions are characterized by a number of overlapping features (cf. Chapter 13.4.2). To clarify what emotion is being expressed by the virtual native speaker, it is necessary to integrate the facial and vocal expressions with each other and with further cues (utterance content, gaze patterns, gestures, postures, and locomotion). Furthermore, it is essential to maintain coherence and consistency (cf. Chapter 7.4 and Chapter 7.5) among the different verbal and non-verbal components of the speaker's emotional displays, and of each display with the content of the speaker's utterance which it accompanies.

By choosing its emotional expressions carefully and portraying them believably, the virtual native speaker can achieve different communicative, social, and task-related goals in its interactions and performances involving the learner and other virtual actors. Some emotional displays serve an identifiable function, e.g. as part of a virtual performance or to encourage the learner to approach the speaker. In contrast, other emotional displays give the impression of being spontaneous reactions of the speaker to an event or situation in the virtual room (although the speaker's displays are always the result of internal deliberation).

There are a number of sources from which the speaker draws information that influences its selection and portrayal of emotional displays:

- The content of the speaker's next utterance;
- The script or plan underlying a virtual performance;
- The speaker's assessment of the current state of affairs in the virtual room;
- Emotional states detected in or inquired from the learner;
- Knowledge of the (communicated) emotional states of other virtual actors;
- The speaker's personality and expressive style;
- The display rules of the culture which the speaker represents.

Considerations of what kinds of emotions should be expressed by the virtual native speaker toward the learner involve a tension between satisfying the hedonic preference of people for positive emotions and the need to make emotional displays appropriate for the current context and the speaker's cultural background (cf. Chapter 13.4.2). To create a comfortable atmosphere in the virtual room, the speaker generally should communicate positive emotions to the learner, exploiting his or her hedonic preference. The intensity of the positive emotions expressed by the speaker may grow as it develops a closer relationship with the learner (cf. Section 16.3.2.4). However, the virtual native speaker should also adapt its expressions when the context of the interaction or performance requires it. For example, the speaker should not show a happy face or speak with a happy voice if another actor hits it with an umbrella or steals an object from its possession.

As an individual, the virtual native speaker should be designed to exhibit idiosyncratic ways of expressing different emotions. Likewise, the other virtual actors as well as the coach and peer agents each require their own distinctive styles of emotion expression. Through regular interactions with the virtual native speaker and the other agents, the learner becomes more and more familiar with their individual expressive styles and improves his or her ability to interpret their emotional displays correctly.

As a representative of its culture, the virtual native speaker should behave in a way that observes the customs and conventions of that culture, which includes the emotional displays of the speaker. When interacting with others, the speaker should choose the appropriate time, place, manner, and addressees for expressing or suppressing different emotions, adhering to the display rules of its culture (cf. Chapter 10.1.3). The resulting authentic emotional displays are instructive for the learner, who can study them and speculate about the display rules of the target culture that underlie the displays observed.

16.3.5.2 Personality

Each fieldwork class has a different virtual native speaker plus a number of other virtual actors. All these agents are individuals in their own right, and each of them should be designed with a personality (cf. Chapter 13.5) that distinguishes this particular agent from other virtual actors and the learner, although similarities may exist and can be exploited to make interactions more effective and enjoyable. The personality not only defines a virtual native speaker as a distinct individual, but it also influences how the speaker interprets and reacts to different situations or events in the virtual room, which may differ from the ways in which other agents perceive and react to the same situations or events.

The personality of a virtual native speaker and its differences and similarities to the personalities of others are reflected in the speaker's appearance (cf. Section 16.3.1) and behaviors (cf. Section 16.3.2), as well as its interactions with the learner (cf. Section 16.3.3) and with other actors during virtual performances (cf. Chapter 16.3.4). The personality is conveyed through all available channels, i.e. both visually and auditorily, which includes characteristic ways in which the agent speaks, bears its body, moves through the virtual room, performs gestures, and maintains eye contact, as well as the speaker's distinctive facial features and movements. Given that various aspects are involved in creating and communicating a personality, it is crucial to maintain coherence (cf. Chapter 7.4) and consistency (cf. Chapter 7.5) among the different components. The personality of a virtual native speaker (or a coach or peer agent) and its expression have to be designed with careful consideration for the necessary level of detail of the individual aspects, but without losing sight of the "big picture" of how these aspects interact.

While distinct rather than uniform personalities for virtual native speakers are the goal, there are certain traits (cf. Table 27 and Table 28) that should be shared by all speakers because they are conducive to both their interactions and relationships with learners. First, a virtual native speaker should be open to new experiences, in particular to meeting new people (learners) from a culture different to its own. This includes a certain degree of curiosity about the learner as a person and as a representative of another culture, but also tolerance for different views and behaviors. Second, the virtual native speaker should be portrayed as an approachable, cooperative, and courteous individual, in order to encourage learners to engage in interactions with and request linguistic data from the speaker. Third, given their important role as learners' informants in fieldwork classes, virtual native speakers have to be dependable individuals. The learner has to know that he or she can rely on the speaker for data and trust the utterances provided by the speaker to be authentic and accurate descriptions of the state of affairs in the virtual room. Fourth, the virtual native speaker should portray an emotionally stable rather than a neurotic persona. In particular, patience is a virtue for virtual native speakers because learners may ask them to repeat the same utterance or performance again and again, or they may repeatedly interrupt and modify performances (cf. Section 16.3.4).

To a certain extent, the virtual native speaker may be able to adapt its personality to the one it detects in the learner, for example along the extroversion-introversion dimension, in order to exploit the positive effects of similarity attraction (people like other people and things better that they perceive as similar to themselves) on the learner's impression of the agent (cf. Chapter 9.2 and Chapter 10.1.1). To assess the learner's personality, the direct approach of asking the learner questions that reveal different personality traits could be used. Alternatively (or in addition), in a future version of the fieldwork environment that permits speech input, the learner's voice characteristics (volume, pitch, pitch range, and speech rate) could be analyzed and mapped to different personality dimensions or types. For example, an extrovert voice exhibits a loud volume level, a high pitch, a wide pitch range, and a fast speech rate (cf. Chapter 13.5).[269] When using a synthetic voice, these parameters can be manipulated to express personality characteristics through the voice that are more similar to those of the learner. Of course, as discussed above, changing the voice personality typically also requires adjustments of an agent's non-verbal behavioral tendencies (e.g. an extrovert voice combined with introvert gesturing would be perceived as inconsistent). While subtle changes in the speaker's personality expression can be made during and across interactions with the learner, more radical adaptations should happen before the first interaction with the learner, since relatively sudden shifts in personality are very likely to be perceived as signs of emotional instability or even a personality disorder.

16.3.5.3 Back-Story

The current virtual native speaker is a relatively 'flat' character, not just with respect to its embodiment (cf. Chapter 16.3.1) but also regarding the story behind the character. All the information that the learner can ever obtain about the speaker is restricted to its name and to a small number of biographical details about the human native speaker who provided the voice talent for the character. In contrast, the enhanced virtual native speaker is designed to be a deeper character with an identity of its own, which includes a distinct back-story that provides significantly more detail than the sketchy vita in the first unit of the fieldwork class.

The back-story of a virtual native speaker requires sufficient breadth and depth to enable the learner to understand what individual (or character) the virtual native speaker is. It integrates the speaker's biographical background with a (growing) collection of scripted experiences, memories, stories, etc. (cf. Chapter 13.6). The contents of the back-story not only allows the learner to get to know the virtual native speaker better as an individual but may also give him or her valuable background information about the culture (through the eyes of one that 'lives' in it). For example, the back-story could include special events in the life of the virtual native speaker, some of which might even happen in the course of the fieldwork class (e.g. birthday, graduation, marriage, death of a family member, etc.). The back-story could illustrate how these events are handled in the culture of the virtual native speaker. By taking an interest in accounts of such events, the learner is getting closer to both the virtual native speaker and its culture.

Back-stories of virtual native speakers (and also of coach agents and peers) should be interesting to learners and help them to relate to the agent as an individual. The back-story of a virtual native speaker may (by accident or by design) exhibit some similarities to the learner's own background, but that is not a requirement. Having said that, it is not necessary to give the character a back-story that oozes with adventure or tragedy either. In fact, such an overdrawn character would be inappropriate for its role in the fieldwork scenario. Real life writes the best stories after all, and portraying a virtual native speaker (or coach or peer) as a 'normal' person, like the learner, makes it easier for the latter to identify with the agent.

The back-story of a particular virtual native speaker should be revealed gradually in the course of the fieldwork class. Letting out the whole story at once might overwhelm the learner and can lead to a loss of interest in the speaker on his or her part because the mystery that was the virtual native speaker was resolved too quickly. Instead, disclosing one aspect at a time helps to maintain the learner's interest in the speaker. Due to the language barrier between the

[269] However, since the virtual native speaker should be perceived as competent, it may be necessary to lower the pitch of its voice (Nass & Brave 2005: 42).

learner and the virtual native speaker, the back-story cannot be told in the language of the fieldwork class but has to be written in English. Learners can read the story or may listen to the narration provided by a coach or peer agent that reveals a particular aspect at the appropriate time.

Each virtual native speaker is a different individual and therefore requires its own back-story. This means that at least one back-story per fieldwork class has to be written for its virtual native speaker. In addition, the other actors that appear in a fieldwork class also each require their own back-story. To manage the amount of writing necessary, several virtual actors may share certain aspects of their backgrounds (due to being close friends or relatives, for example). Furthermore, the back-story of a fieldwork character is not static but progresses in the course of a fieldwork class and from one run of the class to the next because life goes on for these artificial individuals as well. Constantly authoring the life of a single virtual actor, let alone the lives of multiple actors, is not feasible. However, certain changes to the back-story of a virtual native speaker or other agent should be made, at least between two runs of the fieldwork class, to show that the agent's story continues.[270]

16.3.6 Evaluation

Evaluation is an important activity throughout the development life-cycle of software systems, including embodied interactive agents (cf. Chapter 15). In the design and implementation of the virtual native speaker (as well as the coach and the peer),[271] repeated evaluations are essential to generate options for design decisions, guide the choice among design options, assess the effects of the resulting designs, and iteratively improve the overall system. Furthermore, they can help to derive guidelines for the design of future agents for the Virtual Linguistics Campus and other environments, roles, and applications (cf. Chapter 15.1). The following paragraphs discuss some preliminary ideas about what aspects of the virtual native speaker should be evaluated, and how.

From a macro (system-level) evaluation perspective (cf. Chapter 15.3.4), the design of the virtual native speaker can be considered a success if the resulting agent fulfills its functions of informant, performer, individual, cultural representative, and partner of the learner in ways that enhance and facilitate the learner's experience and outcomes of studying a linguistic fieldwork class, both according to objective measures and to the learner's subjective perception. In particular, the virtual native speaker should help the individual learner to develop proficiency in linguistic fieldwork and to construct knowledge about the speaker's language and culture while playing its part in making the time spent by the learner in the virtual room as productive and enjoyable as possible. A successful virtual native speaker serves the learner as a flexible and competent informant that provides relevant and accurate data for the analysis of the target language. In performances with other virtual actors and with the learner, the speaker plays its part believably, comprehensibly, and with its own personal touch. Through all aspects of its appearance and behavior, it presents itself as an appealing individual with a distinct personality

[270] However, from a pragmatic point of view, the question is how likely it is for the same learner to take the same fieldwork class twice. Still, the life stories of virtual native speakers, coach agents, and peer agents should be updated at regular intervals.

[271] While the following discussion only makes reference to the virtual native speaker, it is clear that *all* virtual actors designed for the fieldwork environment have to be evaluated, which includes the coach and peer agents (cf. Section 16.2.3 and Section 16.2.4).

and as an authentic representative of its culture. Finally, in the course of a fieldwork class, learners come to appreciate the speaker as a reliable and trustworthy regular partner in the fieldwork process.

The above criteria for the success of the virtual native speaker as part of the virtual fieldwork environment can be met by developing designs for the individual components of the speaker (embodiment, behavior, conversation, expertise, and emotion and personality) that contribute desired attributes, such as an authentic and inviting appearance; a distinct, intelligible, natural-sounding, and expressive voice; deictically believable combinations of speech, gesture, and locomotion; the ability to produce linguistically and factually accurate verbal descriptions of different states of affairs in the virtual room; and a supportive personality backed up by a deep, interesting back-story. For each design component of the virtual native speaker discussed in the previous sections, a set of questions is provided below as examples of issues that are of interest for evaluating the contribution of that component to the overall success of the speaker (micro evaluation, cf. Chapter 15.3.5):

Embodiment

- Does the speaker's embodiment authentically portray a contemporary representative of the target culture, according to judgments by people from the same culture and from different cultures? Do the speaker's appearance and/or behavior offend either of these groups?
- What are the effects of exploiting cultural stereotypes in the speaker's appearance (and behavior) on the speaker's perception by members of different cultures?
- Do the speaker's physical attributes, clothes, and accessories appeal to learners from the target group? Do they invite learners to interact with the speaker?
- Does the speaker's appearance help learners to understand the speaker's role in the fieldwork environment?
- Does the speaker's appearance support learners' perception of the speaker as an individual that is distinct from other virtual actors?
- What are the effects of the speaker's anthropomorphic embodiment on learners? What expectations with respect to the speaker's characteristics and capabilities does it raise? How does it influence learners' behavior toward the speaker?
- Was the right level of realism chosen for the speaker's embodiment? If not, is the speaker's level of realism too low (below the threshold of believability) or too high (pushing the speaker into the uncanny valley) (cf. Chapter 9.5)?
- Is the quality of the animations created for the speaker acceptable? Are the movements smooth and without perceptible glitches?

Voice

- Does the voice accurately convey the segmental and suprasegmental aspects (i.e. sounds and prosody) of the target language?
- Is the speech sufficiently intelligible for learners to be able to correctly transcribe the speaker's utterances?
- Is it easy and pleasant for learners to listen to the speaker's voice?
- Does the voice have the accent of a native speaker of the target language?
- Is the speaking style (speech rate, volume, pitch, etc.) appropriate?
- How effectively does the voice contribute to the speaker's expressions of different emotions?
- Is the voice chosen perceived as a distinctive characteristic of the speaker as an individual?

- Can the requirements above be satisfied by a synthetic voice or is it necessary to cast a human voice?

Non-Verbal Communication

- Do the speaker's non-verbal communicative behaviors seamlessly integrate with each other and with the speaker's utterances during its performances and interactions with the learner? For example, does the speaker integrate speech, gestures, and locomotion in a natural and unambiguous way when referring to actors, objects, and events in the virtual room?
- Does the speaker follow the conventions of its culture when displaying non-verbal behaviors in its interactions with others? Are potential misinterpretations of behaviors by the learner due to cultural differences cleared up timely and appropriately?
- Is the degree of variability in the speaker's non-verbal behaviors sufficient to prevent the impression of repetitiveness?
- Is the speaker's manner of performing different non-verbal behaviors perceived as consistent with its personality?
- Do modulations of non-verbal behaviors successfully contribute to the communication of the speaker's different emotional, cognitive, and physical states?
- Are the speaker's auditory and visual speech accurately synchronized?

Believability

- What is the relative contribution of human-like communicative behaviors and explicit believability-enhancing behaviors to the speaker's believability as an individual? Which of the two is more important to this end? Do learners notice any difference between a virtual native speaker with and without explicit believability-enhancing behaviors vs. a speaker with and without communicative behaviors?
- Are the speaker's believability-enhancing behaviors sufficiently subdued to be recognizable without being a distraction for learners? Or do they attract too much attention?
- Does the speaker perform its reflexive and idiosyncratic behaviors in a way that is perceived as human-like?
- Does the speaker observe the context of the interaction or performance when carrying out behaviors that enhance its persona?
- Do believability-enhancing behaviors, when displayed in a particular context, cause any misunderstandings regarding the attitudes and intentions of the speaker in that context?
- Does the speaker perform its believability-enhancing behaviors with sufficient variability to avoid the impression of repetitiveness?
- Do learners associate the speaker's idiosyncratic behaviors with its persona? Can learners recognize and remember the speaker better because of these behaviors?

High vs. Low Profile

- Do learners always know where they can find the speaker in the virtual room?
- Do learners know that the speaker is always available for them, even if it is doing something else in the virtual room?
- How much attention does the speaker draw to itself (a) when it provides its services as the learner's informant; and (b) when it is occupied otherwise? Does it attract too much attention in either case?

- How do learners react to the speaker's constant presence? Do they perceive it as comforting or as a source of additional stress? Or are they indifferent to it?
- Does the speaker succeed in maintaining a level of presence in the virtual room that indicates its constant availability without being a source of distraction or disturbance for learners?

Relationship Building

- Do learners perceive the speaker's relationship-building behaviors as intended?
- How do learners react to the speaker's efforts to form a relationship with them?
- If the speaker discloses information about itself, do learners reciprocate in kind?
- What observable changes occur in learners' attitudes and behaviors toward the speaker in the course of a series of virtual fieldwork sessions?
- How do learners view their relationship with the virtual native speaker?
- Do learners accept the speaker as a regular partner in the fieldwork process?
- Do learners think that they can trust and rely on the speaker?

Conversation

- Are the speaker's utterances grammatical?
- Are the utterances delivered with a native pronunciation and prosody (or at least one that comes sufficiently close to it)?
- Do the speaker's vocabulary, constructions, and articulation reflect how the target language is spoken today?
- Do the speaker's utterances provide accurate descriptions of different actors, objects, and events in the virtual room, in the present, past, and future?
- What are the limits of the speaker's descriptive capabilities? What situations in the virtual room is the speaker able or unable to describe with its utterances?
- Are the speaker's utterances appropriate for the context of the interaction or performance in which they occur?
- Do the utterances generated by the virtual native speaker have the right level of complexity for a given learner's knowledge and capabilities?
- How easy (or difficult) is it for learners to elicit the relevant or desired linguistic data from the speaker? Does the graphical (or multimodal) user interface to the virtual room facilitate or hinder this process? Is this interface usable by learners from the target group?
- How accurately can the speaker recognize and interpret the learner's speech and mouse inputs, alone or in combination?
- Does the speaker's dialogue model correctly reflect the development and the current state of the interaction with the learner and/or with other agents?
- Is learner control a successful strategy for the interactions between the learner and the virtual native speaker, or should the speaker take the initiative more often?

Expertise

- Does the speaker's model of the virtual room provide an accurate and constantly up-to-date representation of this dynamically changing environment over time, including previous, future, and hypothetical states and events of the virtual room?

- Is the information in the model sufficiently broad and deep to enable the speaker's language generation component to produce utterances that describe the specified actor, object, or event with accuracy and in appropriate detail?
- Does the ontology created for the virtual room capture the structure of this domain in a way that allows to economically and unambiguously describe the full range of entities and their relationships and facilitates making inferences from the current model of the virtual room?
- Does the use of an external model of the virtual room have the anticipated advantages (cf. Section 16.3.4), or is it necessary to think of an alternative?
- Does the visual authoring tool facilitate the scripting of virtual performances at different levels of detail and complexity that are appropriate for particular groups of users (learners, instructors, and developers)?
- Are the performances scripted using the tool performed by the virtual actors as the user imagined and specified them?
- Do dynamically generated performances illustrate the linguistic (or cultural) concept set as a goal for the improvised scene in a way that is comprehensible and interesting for the individual learner?
- Do both pre-scripted and dynamically generated virtual performances observe restrictions with respect to focus and length? That is, are the performances short and to the point?
- Are interruptions of performances handled smoothly by the virtual actors? Are the break and continuation points chosen appropriately?
- How well does the system deal with learners' modifications to a performance? Are illegal modifications detected and rejected? Are the revised plans for the performance consistent, and do they appropriately reflect the changes requested by the learner?
- In performances with flexible learner participation, are the learner's goals and plans inferred correctly during the performance? Can the virtual actors successfully adapt their respective parts to the learner's actions and intentions?
- Do the virtual performances in general achieve the goal of exposing the learner to more and more varied data from the language of the fieldwork class?
- Do the virtual actors play their parts in a performance in a way that is believable, individual, culturally appropriate, and easy to interpret for the learner?
- Does the information accumulating in the speaker's model of the individual learner enable the speaker to become a better informant, performer, and/or partner in the relationship with the learner?
- Is the learner-specific information obtained from various sources (the learner, external databases, and through observation) accurate and consistent (in itself and with each other)?
- What types of information in the student model (cf. Section 16.3.4) are necessary for the speaker to adapt successfully to the learner? Which ones are superfluous?

Emotion

- Does the speaker's emotion model enable the agent to communicate contextually, culturally, and individually appropriate emotions to the learner and to other agents?
- Is it sufficient for the speaker to deliberately communicate emotions, or does it require an internal simulation of emotional states and mechanisms that condition its emotional responses?
- What emotions should the speaker be able to detect in learners? Which ones should it be able to express itself?

- How accurately can the speaker detect and interpret the spectrum of learners' emotional states that are relevant for its role?
- How does the speaker's failure to confidently identify the learner's emotional state affect its interactions and relationships with learners?
- Are the speaker's emotional displays easy to interpret for learners? Do learners recognize the emotion communicated by the speaker as intended?
- Are the emotional displays of the speaker believable for learners?
- Do learners recognize the speaker's distinctive style of emotion expression?
- Are the emotions expressed appropriate for the current context and addressees? How closely does the speaker adhere to the display rules of its culture when displaying or suppressing emotions to other actors and to the learner?
- Do the speaker's emotional displays exhibit cross-modal consistency? That is, is the emotion communicated through the voice consistent with the one portrayed non-verbally and with the (emotional) content of the speaker's utterances?

Personality

- Is the speaker's personality clearly and consistently communicated through the interplay of its appearance, behavior, and communication?
- Does the speaker portray a personality that is perceived as distinct from the personalities of other agents? Can learners identify the differences and similarities between the speaker's personality and other personalities in the virtual room as well as their own?
- Do learners recognize the speaker's personality as intended, in particular the essential traits openness, approachability, cooperativeness, courtesy, dependability, emotional stability, and patience?
- Do learners think that the speaker's personality is appropriate for its role?

Back-Story

- Does the speaker's back-story provide the appropriate elements and level of detail that allow learners to get to know the (scripted) person behind the virtual native speaker?
- Does the back-story contain elements that appeal to learners and help them to relate to the speaker as an individual?
- What elements of the speaker's back-story draw learners' interest in particular?
- Is the pace at which the speaker's story is revealed in the course of a fieldwork class conducive to maintaining learners' interest in the speaker?
- Is the rate at which the speaker's story is updated sufficient for sustaining the illusion of a progressing life?

Following the principle of participatory design (cf. Chapter 7.13), the development of the virtual native speaker and the other pedagogical agents for the linguistic fieldwork classes should incorporate inputs from all parties that can make contributions to its design, including learners from the target group of the fieldwork classes; experts on linguistic fieldwork; instructional design experts (with knowledge of constructivist methods); experts on usability, affective and social design (cf. Chapter 6); experts on (different design aspects of) embodied interactive software agents; and native speakers of the language of the fieldwork class (who provide expertise on their native language as well as their culture).

Formative evaluations of the virtual native speaker (cf. Chapter 15.3.2) should begin early and recur at regular intervals during the development process to provide guidance at appropriate points (e.g. when casting the voice for the speaker). These evaluations are largely exploratory in nature, generating design options and guiding design choices. They may focus on individual components of the virtual native speaker (see above) but also examine more or less functional prototypes of the virtual room featuring the enhanced speaker.

Preparations for the actual design process involve a substantial amount of desk research (cf. Chapter 15.4.4.4) to collect and review insights from previous work done by others that pertain to the design of embodied interactive software agents in general and of the pedagogical agents for the linguistic fieldwork classes in particular. The results of this desk research were presented in the previous chapters of this thesis. Furthermore, right from the beginning, it is important for developers to have a clear picture of the people that will work with the virtual native speaker. A survey (cf. Chapter 15.4.4.1) in questionnaire format (cf. Chapter 15.4.5.1) could be used to elicit broad user-related information (cf. Chapter 6.1 and Chapter 15.2.2). Longitudinal case studies (cf. Chapter 15.4.4.3) are another useful exploratory tool in the early phases of design. By observing a small number of learners (which may be selected from the respondents of the initial survey) over a series of sessions in the current fieldwork environment, it is possible to obtain detailed descriptive data from which their patterns and strategies of exploring the virtual room and interacting with the current virtual native speaker can be derived. These case studies may also include exploratory interviews with the participants, which elicit their expectations about desirable qualities of a virtual native speaker, influenced by their interactions with the current speaker (cf. Chapter 15.4.5.2).[272] The data obtained from these exploratory studies, while difficult to generalize because the subjects studied are few and selected deliberately rather than randomly, nonetheless provide insights for designing improved facilities for exploration and interaction into the virtual room and the enhanced virtual native speaker.

Experimental user studies (cf. Chapter 15.3.2) are another important instrument for both the formative and the summative evaluations of the virtual native speaker because they allow the design team to assess how representative members of the target group of learners perceive and interact[273] with different versions of the improved speaker agent, including the one that is eventually deployed. Even before a fully functional implementation is available, a series of Wizard-of-Oz experiments with successively refined prototypes of the virtual room and the speaker agent can be conducted to elicit (by way of questionnaires, observations, and think-aloud protocols) how learners behave in their interactions with the virtual native speaker, what learners think of the interactions and the speaker, what confuses them, what they like and dislike, what features they would like to see or change, etc.

Focus groups (cf. Chapter 15.4.5.2) can be employed to broaden the range of voices that can comment on the design of different prototypes of the virtual room and the virtual native speaker. Apart from learners, focus groups may involve experts, instructors, and native speakers

[272] Kim (2007) conducted such interviews with a small group of learners to collect their opinions on desirable characteristics of pedagogical agents as learning companions (PALs), based on their previous experience with such agents.

[273] A problem with many evaluation studies of embodied interactive software agents to date is that they do not involve *interaction* between the subjects and the agents being evaluated (cf. Footnote 66 and Chapter 15.5). In contrast, for determining the success of the design of the virtual native speaker, it is essential to assess the informativeness, pleasantness, and smoothness of learners' interactions with the speaker.

of the target language, who share their attitudes, feelings, impressions, etc. regarding a given prototype presented for discussion. The moderators of the focus groups should not be affiliated with the team designing the virtual native speaker, so they can remain neutral.

Formative evaluations of the virtual native speaker can also benefit from reviews provided by experts on linguistic fieldwork, the language and culture of the fieldwork class, instructional design, and embodied interactive software agents. Expert reviews may be based on heuristics (cf. Chapter 15.3.2), for example when assessing the usability of the enhanced graphical (or multimodal) user interface to the virtual room (cf. Section 16.3.3). For graphical interfaces, established usability principles are already available. A number of heuristics for the design of multimodal systems are indicated as 'implications' in Table 15. Heuristics for the evaluation of dialogue interfaces are discussed by Sanders and Scholtz (2000). The situation is more difficult for the virtual native speaker as such because generally accepted heuristics for the evaluation of embodied interactive software agents do not yet exist. The design principles developed in Chapter 7 are a first step toward developing such heuristics. In general, heuristics for embodied interactive agents have to go beyond concerns for usability and include affective and social aspects (cf. Chapter 6).

Apart from using general heuristics, a collection of domain-specific test cases can be defined by evaluators that cause the virtual native speaker to exhibit certain qualities, such as language generation, emotion expression, and deictically believable communicative behaviors. Test cases for the speaker may include pre-scripted performances; goals for improvised performances; sets of entities, relationships, and events providing topics for the speaker's utterances; and different emotion-generating situations. While pre-defined test cases can be used to systematically test the speaker's capabilities, it is also helpful to give learners in user studies the freedom to spontaneously create situations and performances because this allows evaluators to see how well the speaker handles these unanticipated cases.

When the virtual native speaker is reaching its final development stages, outcome-oriented (summative) evaluations (cf. Chapter 15.3.3) come to the fore. At this stage, experiments (cf. Chapter 15.4.4.2) with subjects representative of the target group of learners should be conducted that evaluate the (almost) finished speaker in its application context (the virtual room in a particular fieldwork class) in terms of quantitative measures of performance and user satisfaction, relative to an appropriate control condition (at the coarsest level, the original virtual room and virtual native speaker) in order to answer specific research questions. These studies should assess the *actual* value (as opposed to the anticipated value) that the enhanced virtual native speaker adds to the virtual fieldwork environment (cf. Chapter 7.1), demonstrate that learners can (or cannot) interact with the speaker and use the enhanced features of the virtual room (authoring of performances, manipulation of the room, speech and multimodal input) successfully to achieve the goals of a linguistic fieldwork class, and identify aspects of the speaker, the room, and the user interface that need to be improved in future versions.

Furthermore, it is important to determine if learners accept the virtual native speaker in its different functions of informant, performer, individual, cultural representative, and partner in the learning process. To determine the long-term effects of the virtual native speaker, its behaviors and components on learners, it is not enough to examine one-time, short interactions between learners and the speaker. Instead, extended evaluation studies have to be conducted that assess how learners' perception of and interaction with the virtual native speaker change over time. These long-term evaluations are conducted with the final, deployed speaker agent in the context of an actual fieldwork class, using real students of that class as test subjects. They can be implemented as case studies documenting the series of interactions between individual learners and the virtual native speaker during a fieldwork class in detail or, at a coarser level, by administering appropriate questionnaires at different points in the course of the class. The data collected from these long-term postproduction evaluations will inform the next development cycle for the virtual native speaker and the virtual room.

16.4 Summary

This chapter worked out a detailed design for the virtual native speaker, an embodied pedagogical software agent which serves as the learner's informant about an unknown language that is explored by the learner in the context and course of a particular linguistic fieldwork class on the Virtual Linguistics Campus. The design extends the existing virtual native speaker role known from the current setup of the fieldwork classes in substantial ways. The enhanced virtual native speaker is no longer just a machine with a face and a voice that plays back recorded utterances at the learner's request but a more flexible, competent, and interesting interaction partner and performer with its own personal and cultural background, which provides various opportunities for the learner to discover aspects of the speaker's language and culture and works to establish a long-term cooperative relationship with the learner.

The design of the virtual native speaker was divided into five components, embodiment, behavior, conversation, expertise, and emotion and personality, to mirror Chapter 9 to Chapter 13 of the thesis that were concerned with these components. Considerations regarding the technical architecture and platform of the virtual native speaker (cf. Chapter 14) were integrated into the description of the speaker's design. A further section discussed preliminary ideas for the evaluation of the enhanced virtual native speaker.

For the embodiment of the virtual native speaker, a full-body, animated 2½D artistic humanoid representation was chosen, whose appearance authentically portrays a distinctive individual representing the target culture. The speaker's appearance helps the learner to quickly understand its identity and role and invites him or her to interact with the speaker.

The virtual native speaker makes use of its embodiment to perform various behaviors that serve different communicative functions during its interactions with the learner and other virtual actors, enhance the speaker's believability as an individual, and facilitate the process of building a relationship with the learner. The communicative behaviors of the virtual native speaker involve the coordinated use of voice, facial expressions, gaze, gesture, posture, and locomotion, whose design was discussed in separate subsections. These behaviors have to be coherent and consistent with each other, as well as appropriate for the current context and the speaker's personality and cultural background. The virtual native speaker achieves believability by behaving in human-like, individual and culturally authentic ways during its interactions with others rather than through explicit believability-enhancing behaviors. The speaker performs a number of reflexive and idiosyncratic behaviors but plays them down so they do not distract the learner. Due to its essential role in the fieldwork environment, the virtual native speaker has to be available all the time, which involves maintaining an appropriate level of presence in the virtual room without drawing unnecessary attention to itself. Relationship building is hindered by the language barrier separating the virtual native speaker and the learner, so several alternative ways for the virtual native speaker to develop its relationship with the learner were described.

The enhanced virtual native speaker is a conversational agent that participates in interactions and performances with other virtual actors and with the learner, which involve short (simulated) dialogues in its native language, in order to present more and more varied linguistic data from that language to the learner. Conversations in the virtual room are restricted both with respect to the domain and the complexity of the utterances produced by the speaker and the other actors. In future versions of the fieldwork environment, learners may be able to use spoken language input to talk to the virtual actors and even to combine speech with mouse actions. Until that time, learners can use an enhanced graphical user interface to the virtual room, whose features were described. More critical than spoken language understanding or multimodal interpretation is the ability of the virtual native speaker to dynamically generate utterances in its native language that describe any state of affairs that is (or was or will be) present or can be created by the learner in the virtual room. The interactions between the virtual native speaker and the human learner are open-ended rather than goal-oriented and do not develop along a predictable path. Still, the virtual native speaker has to maintain both a model of the current interaction and a history of previous interactions. The initiative in learner-speaker interactions lies with the learner, who inquires about events and situations in the virtual room to elicit the corresponding utterances from the speaker. In contrast, virtual performances are characterized by mixed-initiative exchanges among the actors.

The primary expertise of the virtual native speaker lies in its ability to produce authentic and grammatical utterances in its native language that provide accurate and appropriate descriptions of different states of affairs in the virtual room. The speaker requires a dynamically updated model of the virtual room to keep track of changes happening in the room over time as a result of manipulations by the learner, other virtual actors, and the virtual native speaker itself. The possibility of implementing such a model as an external representation consisting of semantic markup written using the languages of the Semantic Web and encoded as annotations to the speaker's environment (the virtual room) was discussed.

The virtual native speaker is not just an informant but also a performer that plays its part in performances involving other virtual actors, objects in the virtual room, and possibly the learner. These performances are interactive experiences for the learner, who can interrupt a performance to query the actors, modify the performance and observe the results, and even play his or her own part in it. Virtual performances can be scripted in advance by developers, instructors, and learners using an authoring tool, or they may be improvised by the actors (possibly including the learner) according to a plan that is either constructed beforehand to illustrate a particular linguistic concept or evolves in response to the learner's involvement. Both pre-scripted and dynamic performances should have a clear focus (on a small set of linguistic concepts) and a restricted duration.

To be able to adapt to the individual learner in the course of a fieldwork class, the virtual native speaker builds and maintains a long-term student model, which accumulates personal details of the learner; his or her preferred input mode; adaptations of the speaker's behavior; the learner's interaction patterns and exploration strategies; what types of situations, events, and performances he or she creates or manipulates; linguistic data and concepts to which the learner has already been exposed; and, finally, the learner's cognitive and emotional states. The speaker obtains information for the student model from the learner or from external sources or by way of making inferences from the learner's observable behavior, although the options for the latter were argued to be limited in the environment of the virtual room.

For the virtual native speaker to be perceived as a sentient and sensitive individual with its own distinctive traits and background, three ingredients were discussed: emotional intelligence, personality, and back-story. First, the speaker requires the ability to handle emotions in order to support its task, persona, and relationship-building efforts. A communication-driven approach to emotion modeling was adopted for the virtual native speaker, given that the simulation of genuine emotions is less important than the need to communicate contextually, culturally, and individually appropriate emotions to the learner. The speaker's options for detecting the learner's emotional state are quite limited given the few input channels available in a web-based environment. In contrast, the virtual native speaker can make use of its face, body, and voice to portray a variety of emotions, both with an identifiable function and (seemingly) spontaneously. The role of the face and the voice in creating the speaker's emotional displays was discussed in some detail. Furthermore, a number of factors that influence the selection and portrayal of emotional displays were identified. Emotional displays of the virtual native speaker should meet the hedonic preference of people for positive emotions while still being appropriate for the current context and the speaker's cultural background. Furthermore, they should be delivered in the speaker's own distinctive style of expression and observe the display rules of the speaker's culture.

The virtual native speaker is endowed with a personality consisting of a set of traits which distinguish the speaker from other individuals and determine how it interprets and reacts to actors, situations, and events in its environment. The personality is communicated through all aspects of the speaker's appearance and behavior, paying attention to detail as well as the "big picture" of the interplay between the characteristics. While each virtual native speaker has its own personality, a number of traits were identified that all speakers should have because of their utility for both their interactions and relationships with the learner. A virtual native speaker should be open to new experiences and people, approachable, cooperative, courteous, dependable, emotionally stable, and patient. On the other hand, the virtual native speaker can also adapt its personality to match the one of the learner. Ways to assess the learner's personality and to change the speaker's personality expression accordingly were described.

The personality of the virtual native speaker is enriched with a detailed back-story that provides biographical details along with scripted experiences, memories, stories, etc. in order to deepen the learner's knowledge of the speaker as an individual and to give the learner insights into the life of a representative of the target culture. The back-story of a virtual native speaker should be interesting but not exaggerated, portraying the speaker as a 'normal' individual that the learner can relate to. In order to sustain the learner's interest, the back-story should be revealed bit by bit in the course of a fieldwork class. Authoring and regularly updating the different back-stories of the virtual native speaker and the other agents involve a lot of work, so strategies for reducing the workload of writers were discussed.

The final section on the design of the virtual native speaker was concerned with its evaluation. The speaker's design (as well as those of the coach, the peer, the enhanced virtual room, the linguistic workbench, the collaborative learning and research environment, and the visual tool for authoring character performances) will be subjected to repeated evaluations that begin early in the development life-cycle, continue throughout, and extend into the deployment of the finished agents and systems. The success of the virtual native speaker is marked by its ability to serve the learner as an informant, performer, individual, cultural representative, and partner in ways that make the learner's work in the virtual fieldwork environment more productive and enjoyable. Extensive sets of detailed questions for the evaluation of the different components of the speaker's design were provided. The remainder of the discussion contributed ideas on how the formative and summative evaluations of the virtual native speaker could be performed, using case studies, desk research, surveys, and experiments.

17 Conclusion

In the dawning age of embodied interactive software agents, computers have moved beyond the old cliché of the cold, dull, and socially inept machine. Instead, they can portray believable, interesting, and sentient social actors that can interact with users in human-like ways and build relationships with them. This thesis investigated the design of embodied interactive software agents as autonomous, adaptive social actors embedded in computer-based environments and equipped with the ability to use their digital embodiments to act on users and their environment through bodily actions, including verbal and non-verbal communicative behaviors. These embodied (inter-) actors can engage users in multimodal dialogues emulating the socially interactive experience of face-to-face communication. In these dialogues, ideally both artificial and human interlocutors can make use of speech, facial expressions, gesture, gaze, posture, and locomotion to regulate the conversation process, contribute content to the ongoing exchange, and convey their emotional and cognitive states. Apart from being effective communicators, embodied interactive agents can attract, engage, and bond with users by portraying themselves as appealing, empathetic, sentient, and deep individuals that exhibit an attractive appearance, believable behaviors, emotional intelligence, an interesting personality, a rich back-story, and a noticeable interest in and persistent efforts to form a relationship with each user.

Embodied interactive software agents can play a variety of roles in different application domains, some of which were described in Chapter 3.2 of this thesis. However, the focus of the present work was on the design of embodied interactive software agents for roles in computer-assisted learning, in particular e-learning (cf. Chapter 2). The potential of these pedagogical software agents lies in their capacity to implement a wide range of instructional interactions (cf. Chapter 12.3) that provide contextualized, qualified, personalized, and timely assistance, cooperation, instruction, motivation, and services for the individual learner as well as groups of learners, without being a distraction or a nuisance to learners (cf. Chapter 4).

To achieve the qualities outlined in the previous paragraphs, it will not do to conceptualize and design an embodied interactive software agent for a given role as anything less than a complete artificial individual that combines communicative, social, and task-related knowledge and skills with an attractive and expressive exterior as well as a rich inner life and background. In essence, the task is to craft a unique character with depth, appeal, and a history and bring it to life as well as teach it to communicate and behave naturally and effectively in its appointed role and toward the user. Many previous efforts to build embodied agents suffered from not having this comprehensive view that, eventually, the whole package (agent) delivered to users matters, rather than just the performance of particular components (e.g. non-verbal behavior, emotion modeling, dialogue capability, etc.), just as little as it is possible to reduce a human being to individual characteristics or capabilities (cf. Chapter 7.6).

A complete agent cannot be effectively designed with the knowledge and techniques of one discipline alone; it is way too complex for that. Neither computer scientists nor artists and animators nor social psychologists can build a successful embodied interactive software agent on their own. Agent developers require expertise from a variety of academic and non-academic sources, including the arts, computer science, linguistics, psychology, pedagogy, robotics, graphics and animation, the social sciences, computer game research and development, the science-fiction literature, (animated) movies, and others. In fact, one important insight of the present research is that to build an embodied interactive software agent which can effectively function in its role, maintain an appealing and believable persona, and establish productive relationships with users, it is necessary to integrate virtually everything that has ever been learned about what is involved in being human, physiologically, psychologically, socially, culturally, and so on, whether inside or outside science. After all, agents are built after as well as for people. Humans provide the positive and negative examples of behaviors, characteristics, and capabilities, according to which embodied interactive agents are modeled, and they will be the partners of these agents in tasks, interactions, and relationships. Therefore, the goal of the present research was to develop a comprehensive, multidisciplinary, and user-oriented view of the design of embodied interactive software agents.

A further concern of the research was to put embodied interactive agent design on a more principled footing. The field is still quite young, and a theory of embodied interactive agents that could provide guidance during design has yet to emerge (although agent designers have heavily borrowed theories and models about human behavior, communication, intelligence, emotion, personality, etc. from other academic fields to inform their work). On the other hand, embodied interactive software agents are complex artifacts, with regard to both their individual components and the interplay of those elements. Their design is an art as well as a craft and a science, involving knowledge and technique as much as intuition and inspiration. There are many opportunities to make less than optimal decisions during the design process, which often impair the perception and/or performance of the final agent. To help designers to avoid a number of pitfalls in agent development, sixteen principles for the design of embodied interactive software agents for education and other domains were proposed in Chapter 7 of the thesis. The summary of the design principles given there is repeated below:

- *Added value*. The agent should possess knowledge and skills (in particular in verbal and non-verbal communication) that induce users to take advantage of the agent's services regularly and over a longer period of time.
- *Perceptible qualities*. The agent should exhibit cues that allow users to perceive that the agent possesses certain desirable qualities (e.g. believability, competence, conversational ability, empathy, and personality).
- *Balanced design*. Designers should avoid mindless maximization of the sophistication of particular components of an agent and instead attempt to keep all aspects of an agent in approximate balance with respect to quality.
- *Coherence*. All behaviors of the agent (speech, gestures, facial expressions, gaze, postures, and locomotion) should fit seamlessly together and into the ongoing flow of the interaction.
- *Consistency*. Agents should not exhibit contradictions within and among their appearance, behaviors, and capabilities but convey the same general persona through all these aspects.
- *Completeness*. A complete agent should integrate all relevant components (including face, body, ears, eyes, brain, tongue, heart, and life) into a coherent and consistent whole.
- *Comprehensibility*. Users should be able to build an accurate interpretation of the agent's actions, beliefs, personality, etc. from the appearance and behavioral cues displayed by the agent.
- *Individuality*. The agent should reveal and express its personal and social identity in every aspect of its being and behavior, which includes emotional dynamics, personality, back-story, idiosyncrasies, and signature behaviors.
- *Variability*. The agent's behavior should not seem repetitive but vary incidentally and unconsciously with respect to the kind, timing, and manner of the behavior being performed.
- *Communicative ability*. The agent should be able to communicate with users and other agents. In human-agent interactions, this involves using all available communication modes (speech, gestures, facial expressions, etc.) in mixed-initiative multimodal conversations with the user.

- *Modularity*. The agent and its components should be designed in a way that facilitates their reuse in other application domains or in other agents. The design should consist of a system of domain-independent and domain-specific modules that are embedded in an overall architecture.
- *Teamwork*. Agent development is a multidisciplinary team effort involving a wide range of human talents that contribute knowledge and techniques ideally covering the entire spectrum of academic and other efforts to understand the human nature and condition.
- *Participatory design*. The agent should be designed with early and continuous involvement of its future users and employers during the development process.
- *Role awareness*. The agent should portray its social role (including its skills, capacities, and limitations) in a way that helps users to understand the role and to appropriately adapt their behavior toward the agent based on its role.
- *Cultural awareness*. The agent should be viewed as both a representative and a product of a particular culture. Its cultural background should be modeled explicitly by (re-) considering the agent's identity, back-story, appearance, content and manner of speech, manner of gesturing, emotional dynamics, social interaction patterns, role, and role dynamics.
- *Relationship building*. The agent should be made relational, i.e. equipped with strategies and behaviors conducive to building and maintaining long-term social-emotional relationships with individual users.

Following these principles, diverse components have to be designed and combined in order to create an embodied interactive software agent which, both in reality and in the user's perception, achieves the three goals of effectively functioning in its appointed social role, believably creating and maintaining an attractive agent persona, and successfully establishing a productive cooperative relationship with the user (cf. Chapter 8.4). A major part of the thesis was devoted to a detailed discussion of the design of the different components of an embodied interactive software agent for education and other areas. These components include:

- The digital representation (*embodiment*) through which the agent appears in and acts on its environment (cf. Chapter 9);
- The *behaviors* performed by the agent to communicate with the user, enhance the believability of its own agent persona, maintain an appropriate level of presence in the environment, and build a relationship with the user (cf. Chapter 10);
- Technologies, components, and resources enabling the agent to participate in *conversations* with the user that involve the use of (spoken) natural language, multiple combined input modes, and embodied communication (cf. Chapter 11);
- The agent's *expertise* consisting of its knowledge of the domain (domain model), the learner (student model), and instructional interactions (instructor model) (cf. Chapter 12);
- Mechanisms and techniques for modeling, interpreting, and expressing *emotions*; specifying and conveying the agent's *personality*; and writing the agent's *back-story* (cf. Chapter 13).

The discussion of each component contributed specific guidelines for its design and involved critical reviews of pertinent theories, concepts, approaches, and technologies. A further chapter dealt with technical aspects of agent design, including different agent architectures, approaches to generating believable behaviors of embodied agents, and hardware and software platforms for software agents (cf. Chapter 14). The importance of evaluation for shaping the process and products of embodied interactive agent development was acknowledged by including a separate chapter on this topic (cf. Chapter 15), which reviewed the host of evaluation techniques described in the literature from the perspective of the evaluation of embodied interactive software agents and further contributed a set of guidelines for the design and implementation of such evaluations.

The first fifteen chapters of the thesis provided an in-depth coverage of the principles and components of embodied interactive software agent design and evaluation. Equipped with the insights from that discussion, Chapter 16 addressed the remaining topic of designing roles for embodied interactive software agents in e-learning. A particular e-learning scenario was chosen as the context for these roles: the linguistic fieldwork classes on the Virtual Linguistics Campus, which provide a constructivist learning environment for linguistic fieldwork in which learners construct knowledge about an unknown language by exploring a virtual room and interacting with a simulated native speaker. Three pedagogical agent roles for the linguistic fieldwork classes were described in Chapter 16:

- The *virtual native speaker* serves as the learner's informant about the unknown language and culture of the fieldwork class. Extending the original virtual native speaker role, this agent is designed to become a more flexible, competent, and interesting interaction partner that is portrayed as an individual with its own personal and cultural background. Instead of delivering pre-recorded utterances on demand, it has the ability to describe any situation in the virtual room about which the learner inquires, including in particular situations that have been created or changed by the learner. The enhanced virtual native speaker is also a performer that can act out pre-scripted or dynamically generated scenes in the virtual room, alone or together with other virtual actors and/or the learner, in order to illustrate particular phenomena of the target language and culture. Furthermore, it makes use of its regular and extended interactions with the learner in the course of a fieldwork class to work toward building a relationship with him or her, which preferably becomes a long-term cooperative partnership.
- The *coach* plays the role of a master teacher in a cognitive apprenticeship relationship with the learner. It models relevant analysis and hypothesis formation skills involved in doing linguistic fieldwork for its apprentice (the learner), monitors and guides the learner while he or she is practicing these skills, and intervenes with problem-solving and motivational support when and where opportune or necessary. The coach's support gradually fades as the learner gains proficiency in accomplishing linguistic fieldwork tasks on his or her own.
- The *peer* is a virtual co-learner whose characteristics, knowledge, and status are comparable to those of the learner. It explores the language and the virtual room together with the learner, providing different kinds of support for social learning in hybrid communities of artificial and human learners. Peer agents can cooperate or compete with the learner in the virtual room, provide companionship during fieldwork sessions, challenge the learner's skills and confidence as troublemakers, serve as a model of best (or worst) practices in linguistic fieldwork, and coach learners on linguistic field methods or be coached by them.

While the coach and peer roles have been documented in the literature (e.g. Chou et al. 2003; Ryokai et al. 2003; Johnson et al. 2004b; McNamara et al. 2004; Baylor & Kim 2005; Biswas et al. 2005; Graesser et al. 2005; Cassell et al. 2007; Tartaro & Cassell 2008), although not for linguistic fieldwork, the virtual native speaker is a new pedagogical agent role that has not yet been described elsewhere.[274] Therefore, Chapter 16 elaborated a design for the virtual native

[274] The Tactical Language Training System (cf. Footnote 261) also provides animated native speakers of foreign languages, but these agents are conversation partners that help learners to practice their

speaker consisting of the five components embodiment, behavior, conversation, expertise, and emotion and personality. Furthermore, some initial ideas for the evaluation of the speaker's design were presented. Insights from the previous chapters were incorporated into the design wherever appropriate. In addition, the discussion illustrated the breadth and depth of the considerations that need to go into the design and evaluation of an embodied interactive pedagogical software agent.

The research presented in this thesis has made contributions to both the theory and practice of designing embodied interactive software agents for roles in e-learning and beyond. However, there are a number of things that remain to be addressed by future research. The ultimate goal is for all three types of pedagogical agents discussed in Chapter 16, the virtual native speaker, the coach, and the peer, to become regular partners of the learner in the linguistic fieldwork classes on the Virtual Linguistics Campus. In other words, the intention is to turn written designs into computer programs that students of the VLC will routinely make use of in their virtual field research. To achieve this ambitious goal, not only the three agent roles and their components have to be developed but also the set of new supporting systems proposed in Chapter 16, including the linguistic workbench (cf. Chapter 16.2.3), the collaborative learning and research environment (cf. Chapter 16.2.4), and the tool for authoring virtual performances (cf. Chapter 16.3.4). Since the resources for building all these agents and systems in parallel are not available, they will be developed one after the other, starting with the virtual native speaker and the authoring tool for the virtual performances, followed by the coach and the linguistic workbench, and finally the peer and the collaborative learning and research environment. Having said that, it is also possible to assign a separate team, possibly at a different institution, to the development of each agent-system combination.

While the supporting systems can be used in all fieldwork classes, a different virtual native speaker has to be created for each fieldwork class, plus a number of virtual actors playing in the performances. Furthermore, several different coach and peer agents are required. In order to reduce the amount of development, at first, one virtual native speaker, one coach, and one peer agent will be built, one by one, for a single linguistic fieldwork class. Once the designs of the different agents are mature, it will be necessary to think about how they can be recreated for other fieldwork classes (or within the same fieldwork class) without producing identical twins or cheap copies, i.e. how to simultaneously achieve economy and individuality in embodied interactive agent design.

Each agent and system will be developed iteratively by building and evaluating a series of successively refined prototypes (cf. Chapter 15.3.1). Real students of the VLC will use these prototypes, both in controlled contexts and as part of their actual work in the linguistic fieldwork class. For the time being, learners will interact with the agents via a graphical user interface; however, speech and multimodal input are options for future versions of the virtual linguistic fieldwork environment (cf. Chapter 16.3.3).

True to the philosophy of the Virtual Linguistics Campus "What has been built for the Campus, becomes a part of the Campus," the virtual native speaker, the coach, and the peer will be developed with a view to giving them a permanent employment in the linguistic fieldwork classes. As a result, some day in the (hopefully not too distant) future, the Virtual Linguistics Campus will have a number of new members that are not made of flesh and blood but of bits and bytes. An interesting area of research will be to investigate how the humans on the Campus

communication skills in the target language, rather than informants from which learners elicit data for linguistic analysis.

will respond to the virtual native speaker, the coach, the peer, and a host of other agents that do not even have a name yet, including how their responses to these agents will change over time once novelty gives way to routine. Do the virtual native speaker, the coach, and the peer, each in its own particular way, add actual value to the learner's experience in the linguistic fieldwork classes on a continuing basis? Do learners perceive the qualities and capabilities of these agents and make use of them as intended? Do they still enjoy interacting with the agents after several weeks into a fieldwork class, or have the interactions and performances become dull and predictable? Do learners accept each agent as a partner in the learning process? Do they treat the agents well or abuse them? These are just some of the questions that will have to be addressed by long-term studies of the different agents in their respective roles and environments (cf. Chapter 16.3.6).

Deploying any piece of software in the real world inevitably involves the risk of failure. The software may not be what people need; it may be unusable or boring. Through careful design and evaluation, with early and continuous involvement of the target audience, the embodied interactive pedagogical software agents on the VLC may be spared the fate of rejection suffered by previous unsuccessful agent designs, such as those analyzed in Chapter 8. It is the goal of the author's work and his hope that the users of the Virtual Linguistics Campus will come to accept and appreciate the different pedagogical agents described in this thesis. If learners value the services provided by each of these agents, acknowledge its individuality, make it their partner in learning, and look forward to the next session with the agent, then the agent's design can be called a success. Anything less would have to be considered a failure.

Eben-Ezer.

References

Aamodt, Agnar / Plaza, Enric. 1994. "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches." In: AI Communications. 7(1). March 1994. pp. 39–59.

Aarseth, Espen J. 1997. Cybertext: Perspectives on Ergodic Literature. Baltimore, Maryland: John Hopkins University Press.

Abbattista, Fabio / Catucci, Graziano / Semeraro, Gianni / Zambetta, Fabio. 2004. "SAMIR: A Smart 3D Assistant on the Web." In: PsychNology Journal. 2(1). 2004. pp. 43–60. Available Online. URL: http://www.psychnology.org/File/PSYCHNOLOGY_JOURNAL_2_1_ABBATTISTA.pdf. Last Access: 5 September 2009.

Abdullah, Sophiana C. 2003. "Student Modelling by Adaptive Testing: A Knowledge-Based Approach." Ph.D. Thesis. School of Computing, University of Kent. Canterbury, United Kingdom. June 2003. Available Online. URL: http://www.cs.kent.ac.uk/pubs/2003/1719/content.pdf. Last Access: 4 September 2009.

Abry, Christian / Lallouache, Mohamed-Tahar / Cathiard, Marie-Agnés. 1996. "How Can Coarticulation Models Account for Speech Sensitivity to Audio-Visual Desynchronization?." In: Stork & Hennecke. (eds.). 1996. pp. 247–255.

Abu Shawar, Bayan / Atwell, Eric. 2007. "Chatbots: Are They Really Useful?." In: LDV-Forum. 22(1). 2007. pp. 29–49.

Achard, Michel / Kemmer, Suzanne. (eds.). 2004. Language, Culture, and Mind. Stanford, California: CSLI Publications.

Adolphs, Ralph. 2005. "Could a Robot Have Emotions?: Theoretical Perspectives from Social Cognitive Neuroscience." In: Fellous & Arbib. (eds.). 2005. pp. 9–25.

Agre, Philip E. / Chapman, David. 1987. "Pengi: An Implementation of a Theory of Activity." In: Proceedings of the 6th National Conference on Artificial Intelligence (AAAI '87). Seattle, Washington, USA. 13–17 July 1987.

Aïmeur, Esma / Frasson, Claude. 1996. "Analyzing a New Learning Strategy According to Different Knowledge Levels." In: Computers in Education. 27(2). September 1996. pp. 115–127.

Alba, Joseph / Lynch, John / Weitz, Barton / Janiszewksi, Chris / Lutz, Richard / Sawyer, Alan / Wood, Stacy. 1997. "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces." In: Journal of Marketing. 61(3). July 1997. pp. 38–53.

Alexander, Bryan. 2004. "Going Nomadic: Mobile Learning in Higher Education." In: EDUCAUSE Review. 39(5). September–October 2004. pp. 28–35. Available Online. URL: http://net.educause.edu/ir/library/pdf/ERM0451.pdf. Last Access: 5 September 2009.

Allbeck, Jan M. / Badler, Norman I. 2004. "Creating Embodied Agents with Cultural Context." In: Payr & Trappl. (eds.). 2004. pp. 107–125.

Allen, James F. 1994. Natural Language Understanding (2nd ed.). Amsterdam: Benjamin Cummings.

Allen, James F. 2001. "Natural Language Processing." In: Wilson & Keil. (eds.). 2001. pp. 592–594.

Allen, James F. / Byron, Donna K. / Dzikovska, Myroslava / Ferguson, George / Galescu, Lucian / Stent, Amanda. 2001a. "Towards Conversational Human-Computer Interaction." In: AI Magazine. 22(4). Winter 2001. pp. 27–38.

Allen, James F. / Ferguson, George / Stent, Amanda. 2001b. "An Architecture for More Realistic Conversational Systems." In: Proceedings of the 6th International Conference on Intelligent User Interfaces 2001 (IUI 2001). Santa Fe, New Mexico, USA. 14–17 January 2001.

Alpaydin, Ethem. 2004. Introduction to Machine Learning. Cambridge, Massachusetts: The MIT Press.

Alshawi, Hiyan. (ed.). 1992. The Core Language Engine. Cambridge, Massachusetts: The MIT Press.

Alter, Kai / Pirker, Hannes / Finkler, Wolfgang. (eds.). 1997. Concept to Speech Generation Systems. Proceedings of a Workshop in Conjunction with the 35th Annual Meeting of the Association for Computational Linguistics. Universidad Nacional de Educación a Distancia. Madrid, Spain. 11 July 1997.

Alvarez-Torres, Maria J. / Mishra, Punyashloke / Zhao, Yong. 2001. "Judging a Book by Its Cover!: Cultural Stereotyping of Interactive Media and Its Effect on the Recall of Text Information." In: Journal of Educational Multimedia and Hypermedia. 10(2). June 2001. pp. 161–183.

Amižić, Ani / Stankov, Slavomir / Rosić, Marko. 2002. "Model Tracing: A Diagnostic Technique in Intelligent Tutoring Systems." In: Proceedings of the Central European Exchange Program for University Studies (CEEPUS) Summer School on Modern Methods in Control. Split, Croatia. 16–30 June 2002.

Andersen, Janis F. 1979. "Teacher Immediacy as a Predictor of Teaching Effectiveness." In: Nimmo. (ed.). 1979. pp. 543–559.

Anderson, John R. 1983. The Architecture of Cognition. Cambridge, Massachusetts: Harvard University Press.

Anderson, John R. 1993. Rules of the Mind. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Anderson, John R. / Bothell, Daniel / Byrne, Michael D. / Douglass, Scott / Lebiere, Christian / Qin, Yulin. 2004. "An Integrated Theory of the Mind." In: Psychological Review. 111(4). October 2004. pp. 1036–1060.

Anderson, John R. / Lebiere, Christian. (eds.). 1998. The Atomic Components of Thought. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Anderson, John R. / Corbett, Albert T. / Koedinger, Kenneth R. / Pelletier, Ray. 1995. "Cognitive Tutors: Lessons Learned." In: The Journal of the Learning Sciences. 4(2). 1995. pp. 167–207.

Anderson, Terry D. 2003a. "Modes of Interaction in Distance Education: Recent Developments and Research Questions." In: Moore & Anderson. (eds.). 2003. pp. 129–144.

Anderson, Terry D. 2003b. "Getting the Mix Right Again: An Updated and Theoretical Rationale for Interaction." In: The International Review of Research in Open and Distance Learning. 4(2). October 2003. Available Online. URL: http://www.irrodl.org/index.php/irrodl/article/view/149/708. Last Access: 5 September 2009.

Anderson, Terry D. 2004. "Toward a Theory of Online Learning." In: Anderson & Elloumi. (eds.). 2004. pp. 33–60.

Anderson, Terry D. / Elloumi, Fathi. (eds.). 2004. Theory and Practice of Online Learning. Athabasca, Alberta, Canada: Athabasca University. Available Online. URL: http://cde.athabascau.ca/online_book/pdf/TPOL_book.pdf. Last Access: 5 September 2009.

Anderson, Terry D. / Garrison, D. Randy. 1998. "Learning in a Networked World: New Roles and Responsibilities." In: Gibson. (ed.). 1998a. pp. 97–112.

Anderson, Terry D. / Whitelock, Denise. 2004. “The Educational Semantic Web: Visioning and Practicing the Future of Education.” In: Journal of Interactive Media in Education. Special Issue on the Educational Semantic Web. 2004(1). 21 May 2004. Available Online. URL: http://www-jime.open.ac.uk/2004/1/editorial-2004-1.pdf. Last Access: 5 September 2009.

André, Elisabeth / Klesen, Martin / Gebhard, Patrick / Allen, Steve / Rist, Thomas. 1999a. “Integrating Models of Personality and Emotions into Lifelike Characters.” In: Proceedings of the Workshop on Affective Interactions: Towards a New Generation of Interfaces. 3rd i3 Annual Conference (i3AC ’99). Siena, Italy. 20–22 October 1999.

André, Elisabeth / Rist, Thomas / Müller, Jochen. 1999b. “Employing AI Methods to Control the Behavior of Animated Interface Agents.” In: Applied Artificial Intelligence. 13(4–5). May 1999. pp. 415–448.

Androutsopoulos, Ion / Aretoulaki, Maria. 2003. “Natural Language Interaction.” In: Mitkov. (ed.). 2003. pp. 629–649.

Ankolekar, Anupriya / Krötzsch, Markus / Tran, Thanh / Vrandečić, Denny. 2008. “The Two Cultures: Mashing up Web 2.0 and the Semantic Web.” In: Journal of Web Semantics: Science, Services and Agents on the World Wide Web. 6(1). February 2008. pp. 70–75.

Antoniou, Grigoris / van Harmelen, Frank. 2004. A Semantic Web Primer. Cambridge, Massachusetts: The MIT Press.

Antoniou, Grigoris / van Harmelen, Frank. 2008. A Semantic Web Primer (2nd ed.). Cambridge, Massachusetts: The MIT Press.

Antoniou, Grigoris / Skylogiannis, Thomas / Bikakis, Antonis / Doerr, Martin / Bassiliades, Nick. 2007. “DR-BROKERING: A Semantic Brokering System.” In: Knowledge-Based Systems. 20(1). February 2007. pp. 61–72.

Aoki, Kumiko. 2000. “Taxonomy of Interactivity on the Web.” Paper Presented at Internet Research 1.0. 1st Conference of the Association of Internet Researchers (AoIR 2000). Lawrence, Kansas, USA. 14–17 September 2000.

Arbaugh, J. Ben. 2000. “Virtual Classroom Characteristics and Student Satisfaction with Internet-Based MBA Courses.” In: Journal of Management Education. 24(1). February 2000. pp. 32–54.

Arbib, Michael A. (ed.). 2003. The Handbook of Brain Theory and Neural Networks (2nd ed.). Cambridge, Massachusetts: The MIT Press.

Argyle, Michael / Cook, Mark. 1976. Gaze and Mutual Gaze. Cambridge: Cambridge University Press.

Argyle, Michael. 1988. Bodily Communication (2nd ed.). London: Methuen.

Arnold, Douglas J. / Balkan, Lorna / Humphreys, R. Lee / Meijer, Siety / Sadler, Louisa. 1994. Machine Translation: An Introductory Guide. Oxford: Blackwell Publishing.

Arnold, John. 2003. “Emotion Modeling with Synthetic Agents.” Project Report. Graduate Course CMPUT 603. Department of Computing Science, University of Alberta. Edmonton, Alberta, Canada. 2 December 2003. Available Online. URL: http://www.islandofdoom.com/projects/jarnold603report.pdf. Last Access: 5 September 2009.

Arnold, Patricia. 2005. “Einsatz Digitaler Medien in der Hochschullehre aus Lerntheoretischer Sicht.” E-teaching.org, Portalbereich Didaktisches Design. 12 January 2005. Available Online. URL: http://www.e-teaching.org/didaktik/theorie/lerntheorie/arnold.pdf. Last Access: 5 September 2009.

Arppe, Antti / Carlson, Lauri / Lindén, Krister / Piitulainen, Jussi / Suominen, Mickael / Vainio, Martti / Westerlund, Hanna / Yli-Jyrä, Anssi. (eds.). 2005. Inquiries into Words, Constraints and Contexts.

Festschrift for Kimmo Koskenniemi on His 60th Birthday. Stanford, California: CSLI Publications. Available Online. URL: http://csli-publications.stanford.edu/koskenniemi-festschrift/kk-festschrift-all-2005.pdf. Last Access: 5 September 2009.

Asimov, Isaac. 1954, 1997. The Caves of Steel. Paperback Edition. London: Harper Collins Publishers.

Asimov, Isaac. 1957, 1991. The Naked Sun. Paperback Edition. New York: Bantam Books.

Asimov, Isaac. 1983, 1994. The Robots of Dawn. Paperback Edition. New York: Bantam Books.

Asimov, Isaac. 1985, 1996. Robots and Empire. Paperback Edition. London: Harper Collins Publishers.

Atkinson, Robert K. 2002. "Optimizing Learning from Examples Using Animated Pedagogical Agents." In: Journal of Educational Psychology. 94(2). June 2002. pp. 416–427.

Atkinson, Robert K. / Mayer, Richard E. / Merrill, Mary M. 2005. "Fostering Social Agency in Multimedia Learning: Examining the Impact of an Animated Agent's Voice." In: Contemporary Educational Psychology. 30(1). January 2005. pp. 117–139.

Attardo, Salvatore. 1994. Linguistic Theories of Humor. Berlin: Mouton de Gruyter.

Atwell, Eric. 1999. The Language Machine. London: The British Council. Available Online. URL: http://www.comp.leeds.ac.uk/eric/atwell99.pdf. Last Access: 5 September 2009.

Atwell, Eric. 2005. "Web Chatbots: The Next Generation of Speech Systems?." In: EUROPEAN CEO Journal. November–December 2005. pp. 142–143.

Aust, Harald / Schröer, Olaf. 1998. "An Overview of the Philips Dialog System." In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop. Lansdowne, Virginia, USA. 8–11 February 1998.

Austin, John L. 1962. How to Do Things with Words. Cambridge, Massachusetts: Harvard University Press.

Aycock, Alan / Garnham, Carla / Kaleta, Robert. 2002. "Lessons Learned from the Hybrid Course Project." In: Teaching with Technology Today. 8(6). 20 March 2002. Available Online. URL: http://www.uwsa.edu/ttt/articles/garnham2.htm. Last Access: 5 September 2009.

Aylett, Ruth S. 1999. "Narrative in Virtual Environments: Towards Emergent Narrative." In: Proceedings of the AAAI Fall Symposium on Narrative Intelligence. North Falmouth, Massachusetts, USA. 5–7 November 1999.

Aylett, Ruth S. / Paiva, Ana / Woods, Sarah / Hall, Lynne / Zoll, Carsten. 2008. "Expressive Characters in Anti-Bullying Education." In: Cañamero & Aylett. (eds.). 2008. pp. 161–176.

Ayres, Paul / Sweller, John. 2005. "The Split-Attention Principle in Multimedia Learning." In: Mayer. (ed.). 2005a. pp. 135–146.

Babu, Sabarish V. / Suma, Evan / Barnes, Tiffany / Hodges, Larry F. 2007. "Can Immersive Virtual Humans Teach Social Conversational Protocols?." In: Proceedings of the Annual IEEE Virtual Reality Conference (VR 2007). Charlotte, North Carolina, USA. 10–14 March 2007.

Backus, John W. 1959. "The Syntax and Semantics of the Proposed International Algebraic Language of the Zurich ACM-GAMM Conference." In: Proceedings of the International Conference on Information Processing. Paris, France. 15–20 June 1959.

Badler, Norman I. 1997. "Virtual Humans for Animation, Ergonomics, and Simulation." In: Proceedings of the Workshop on Non-Rigid and Articulated Motion. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97). San Juan, Puerto Rico. 17–19 June 1997.

Badler, Norman I. / Allbeck, Jan M. 2001. “Towards Behavioral Consistency in Animated Agents.” In: Magnenat-Thalmann & Thalmann. (eds.). 2001. pp. 191–205.

Badler, Norman I. / Bindiganavale, Rama / Allbeck, Jan M. / Schuler, William / Zhao, Liwei / Palmer, Martha. 2000. “Parameterized Action Representation for Virtual Human Agents.” In: Cassell et al. (eds.). 2000a. pp. 256–284.

Badler, Norman I. / Philipps, Cary B. / Webber, Bonnie L. 1993. Simulating Humans: Computer Graphics, Animation, and Control. New York: Oxford University Press.

Baeza-Yates, Ricardo A. / Ribeiro-Neto, Berthier 1999. Modern Information Retrieval. Boston, Massachusetts: Addison-Wesley.

Baffes, Paul T. 1994. “Automatic Student Modeling and Bug Library Construction Using Theory Refinement.” Ph.D. Thesis. The University of Texas at Austin. Austin, Texas, USA. December 1994. Available Online. URL: http://www.cs.utexas.edu/users/ml/papers/assert-dissertation-94.pdf. Last Access: 5 September 2009.

Baffes, Paul T. / Mooney, Raymond. 1996. “Refinement-Based Student Modeling and Automated Bug Library Construction.” In: International Journal of Artificial Intelligence in Education. 7(1). 1996. pp. 75–116.

Bailenson, Jeremy N. / Blascovich, Jim. 2004. “Avatars.” In: Bainbridge. (ed.). 2004. Volume 1. pp. 64–68.

Bailenson, Jeremy N. / Swinth, Kim / Hoyt, Crystal / Persky, Susan / Dimov, Alex / Blascovich, Jim. 2005. “The Independent and Interactive Effects of Embodied Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments.” In: Presence: Teleoperators and Virtual Environments. 14(4). August 2005. pp. 379–393.

Bainbridge, William S. (ed.). 2004. Berkshire Encyclopedia of Human-Computer Interaction. 2 Volumes. Great Barrington, Massachusetts: Berkshire Publishing Group LLC.

Ball, Eugene. 2003. “A Bayesian Heart: Computer Recognition and Simulation of Emotion.” In: Trappl et al. (eds.). 2003. pp. 303–332.

Ball, Eugene / Breese, Jack. 2000. “Emotion and Personality in a Conversational Agent.” In: Cassell et al. (eds.). 2000a. pp. 189–219.

Ball, Eugene / Ling, Dan / Kurlander, David / Miller, John / Pugh, David / Skelly, Tim / Stankosky, Andy / Thiel, David / van Dantzich, Maarten / Wax, Trace. 1997. “Lifelike Computer Characters: The Persona Project at Microsoft Research.” In: Bradshaw. (ed.). 1997. pp. 191–222.

Bandura, Albert. 1977. Social Learning Theory. Englewood Cliffs, New Jersey: Prentice Hall.

Bannan-Ritland, Brenda. 2002. “Computer-Mediated Communication, eLearning, and Interactivity: A Review of the Research.” In: Quarterly Review of Distance Education. 3(2). Summer 2002. pp. 161–179.

Banse, Rainer / Scherer, Klaus R. 1996. “Acoustic Profiles in Vocal Emotion Expression.” In: Journal of Personality and Social Psychology. 70(3). March 1996. pp. 614–636.

Barker, Philip. 1994. “Designing Interactive Learning.” In: de Jong & Sarti. (eds.). 1994. pp. 1–30.

Barker, Tim. 2006. “Ethical Dimensions of Synthetic Personalities.” In: Ruttkay et al. (eds.). 2006. p. 6.

Barnouw, Erik. (ed.). 1989. International Encyclopedia of Communications. 4 Volumes. New York: Oxford University Press.

Barr, Avron / Feigenbaum, Richard A. (eds.). 1981. The Handbook of Artificial Intelligence. Volume 1. Los Altos, California: William Kaufmann.

Barros, Flávia A. / Robin, Jacques. 1997. "Processamento de Linguagem Natural." Course Notes. Departamento de Informática, Universidade Federal de Pernambuco. Recife, Brazil. March 1997. No Longer Available Online. URL: http://www.inf.unisinos.br/~renata/cursos/pln/pln-apostila2.pdf. Last Access: 30 October 2007.

Barsoux, Jean-Louis. 1993. Funny Business: Humour, Management and Business Culture. London: Cassell.

Bartle, Richard A. 1990. "Interactive Multi-User Computer Games." Technical Report. MUSE Ltd., British Telecom plc. December 1990. Available Online. URL: ftp://ftp.lambda.moo.mud.org/pub/MOO/papers/mudreport.txt. Last Access: 5 September 2009.

Bartle, Richard A. 2003. Designing Virtual Worlds. Indianapolis, Indiana: New Riders Publishing.

Bartlett, Marian S. / Movellan, Javier R. / Littlewort, Gwen / Braathen, Bjorn / Frank, Mark G. / Sejnowski, Terrance J. 2005. "Towards Automatic Recognition of Spontaneous Facial Actions." In: Ekman & Rosenberg. (eds.). 2005. pp. 393–412.

Bartneck, Christoph / Rosalia, Chioke / Menges, Rutger / Deckers, Inèz. 2005. "Robot Abuse: A Limitation of the Media Equation." In: Proceedings of the Workshop on Agent Abuse. 10th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2005). Rome, Italy. 12–16 September 2005.

Bateman, John A. 2002. "Natural Language Generation: An Introduction and Open-Ended Review of the State of the Art." Faculty 10 (Language and Literary Studies), University of Bremen. Bremen, Germany. September 2002. Available Online. URL: http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/info-pages/nlg/ATG01/ATG01.html. Last Access: 4 September 2009.

Bateman, John A. / Zock, Michael. 2003. "Natural Language Generation." In: Mitkov. (ed.). 2003. pp. 284–304.

Bates, Joseph. 1992. "The Nature of Characters in Interactive Worlds and the Oz Project." Technical Report CMU-CS-92-200. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. October 1992. Available Online. URL: http://www.cs.cmu.edu/afs/cs/project/oz/web/papers/loeffler.ps. Last Access: 5 September 2009.

Bates, Joseph. 1994. "The Role of Emotion in Believable Agents." In: Communications of the ACM. 37(7). July 1994. pp. 122–125.

Bates, Joseph / Loyall, A. Bryan / Reilly, W. Scott. 1992. "Integrating Reactivity, Goals, and Emotion in a Broad Agent." Technical Report CMU-CS-92-142. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. May 1992. Available Online. URL: http://www.cs.cmu.edu/afs/cs/project/oz/web/papers/CMU-CS-92-142.ps. Last Access: 5 September 2009.

Baylor, Amy L. 1999. "Intelligent Agents as Cognitive Tools for Education." In: Educational Technology. 39(2). March–April 1999. pp. 36–40.

Baylor, Amy L. 2001. "Permutations of Control: Cognitive Considerations for Agent-Based Learning Environments." In: Journal of Interactive Learning Research. 12(4). October 2001. pp. 403–425.

Baylor, Amy L. 2005. "The Impact of Pedagogical Agent Image on Affective Outcomes." In: Proceedings of the Workshop "Affective Interactions: The Computer in the Affective Loop." 10th International Conference on Intelligent User Interfaces (IUI 2005). San Diego, California, USA. 9–12 January 2005.

Baylor, Amy L. / Ebbers, Suzanne. 2003. "The Pedagogical Agent Split-Persona Effect: When Two Agents Are Better than One." In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2003). Honolulu, Hawaii, USA. 23–28 June 2003.

Baylor, Amy L. / Kim, Yanghee. 2003a. "Validating Pedagogical Agent Roles: Expert, Motivator, and Mentor." In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2003). Honolulu, Hawaii, USA. 23–28 June 2003.

Baylor, Amy L. / Kim, Yanghee. 2003b. "The Role of Gender and Ethnicity in Pedagogical Agent Perception." In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2003). Phoenix, Arizona, USA. 7–11 November 2003.

Baylor, Amy L. / Kim, Yanghee. 2004. "Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role." In: Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS 2004). Maceió, Alagoas, Brazil. 30 August–3 September 2004.

Baylor, Amy L. / Kim, Yanghee. 2005. "Simulating Instructional Roles Through Pedagogical Agents." In: International Journal of Artificial Intelligence in Education. 15(2). 2005. pp. 95–115.

Baylor, Amy L. / Plant, E. Ashby. 2005. "Pedagogical Agents as Social Models for Engineering: The Influence of Agent Appearance on Female Choice." In: Proceedings of the 12th International Conference on Artificial Intelligence in Education (AI-ED 2005). Amsterdam, The Netherlands. 18–22 July 2005.

Baylor, Amy L. / Rosenberg-Kima, Rinat B. 2006. "Interface Agents to Alleviate Online Frustration." In: Proceedings of the 7th International Conference on the Learning Sciences (ICLS 2006). Bloomington, Indiana, USA. 27 June–1 July 2006.

Baylor, Amy L. / Rosenberg-Kima, Rinat B. / Plant, E. Ashby. 2006. "Interface Agents as Social Models: The Impact of Appearance on Females' Attitude Toward Engineering." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2006). Montreal, Quebec, Canada. 22–27 April 2006.

Baylor, Amy L. / Ryu, Jeeheon. 2003. "Does the Presence of Image and Animation Enhance Pedagogical Agent Persona?." In: Journal of Educational Computing Research. 28(4). 2003. pp. 373–394.

Baylor, Amy L. / Shen, E / Huang, Xiaoxia. 2003. "Which Pedagogical Agent do Learners Choose? The Effects of Gender and Ethnicity." In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2003). Phoenix, Arizona, USA. 7–11 November 2003.

Beck, Joseph / Stern, Mia / Haugsjaa, Erik. 1996. "Applications of AI in Education." In: ACM Crossroads. 3(1). Fall 1996. Available Online. URL: http://www.acm.org/crossroads/xrds3-1/aied.html. Last Access: 5 September 2009.

Beck, Joseph / Stern, Mia / Woolf, Beverly P. 1997. "Cooperative Student Models." In: Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED '97). Kobe, Japan. 18–22 August 1997.

Bednar, Anne K. / Cunningham, Donald / Duffy, Thomas M. / Perry, J. David. 1992. "Theory into Practice: How do We Link?." In: Duffy & Jonassen. (eds.). 1992. pp. 17–34.

Beesley, Kenneth R. / Karttunen, Lauri. 2003. Finite-State Morphology. Stanford, California: CSLI Publications.

Bekey, George A. 2005. Autonomous Robots: From Biological Inspiration to Implementation and Control. Cambridge, Massachusetts: The MIT Press.

Bell, Lori / Trueman, Rhonda B. (eds.). 2008. Virtual Worlds, Real Libraries: Librarians and Educators in Second Life and Other Multi-User Virtual Environments. Medford, New Jersey: Information Today, Inc.

Beller, Sieghard. 2004. Empirisch Forschen Lernen: Konzepte, Methoden, Fallbeispiele, Tipps. Bern: Huber Verlag.

Ben-Gal, Irad E. 2007. “Bayesian Networks.” In: Ruggeri et al. (eds.). 2007. pp. 179–184.

Benoît, Christian. 2000. “The Intrinsic Bimodality of Speech Communication and the Synthesis of Talking Faces.” In: Taylor et al. (eds.). 2000. pp. 485–502.

Bente, Gary / Krämer, Nicole C. / Trogemann, Georg / Piesk, Jens / Fischer, Oliver. 2000. “Conversing with Electronic Devices: An Integrated Approach towards the Integration and Evaluation of Nonverbal Behavior in Face-to-Face Like Interface Agents.” In: Proceedings of the International Workshop on Intelligent Interactive Assistance and Mobile Multimedia Computing (IMC 2000). Rostock-Warnemünde, Germany. 9–10 November 2000.

Bereiter, Carl. 1994. “Implications of Postmodernism for Science, or, Science as Progressive Discourse.” In: Educational Psychologist. 29(1). January 1994. pp. 3–12.

Berge, Zane L. 1999. “Interaction in Post-Secondary Web-Based Learning.” In: Educational Technology. 39(1). January–February 1999. pp. 5–11.

Bergen, Benjamin / Binsted, Kim. 2004. “The Cognitive Linguistics of Scalar Humor.” In: Achard & Kemmer. (eds.). 2004. pp. 79–92.

Berners-Lee, Tim. 2000. Weaving The Web. London: Texere Publishing.

Berners-Lee, Tim / Fielding, Roy T. / Masinter, Larry. 2005. “Uniform Resource Identifier (URI): Generic Syntax.” Network Working Group. RFC 3986. January 2005. Available Online. URL: http://gbiv.com/protocols/uri/rfc/rfc3986.html. Last Access: 5 September 2009.

Berners-Lee, Tim / Hendler, James A. / Lassila, Ora. 2001. “The Semantic Web.” In: Scientific American. 284(5). May 2001. pp. 34–43.

Berry, Dianne C. / Butler, Laurie T. / de Rosis, Fiorella. 2005. “Evaluating a Realistic Agent in an Advice-Giving Task.” In: International Journal of Human-Computer Studies. 63(3). September 2005. pp. 304–327.

Bersin, Josh. 2004. The Blended Learning Book: Best Practices, Proven Methodologies, and Lessons Learned. San Francisco: Pfeiffer.

Beskow, Jonas / Cassell, Justine / Heijlen, Dirk / Krahmer, Emiel / Marriott, Andrew / Massaro, Dominic / Noot, Han / Olivier, Patrick / Pele, Danielle. 2004. “Micro-Evaluation.” Final Preliminary Report. Working Group 3. Dagstuhl Seminar 04121 on Evaluating Embodied Conversational Agents. Castle Dagstuhl, Wadern, Germany. 15–19 March 2004. Available Online. URL: http://wwwhome.cs.utwente.nl/~zsofi/eeca/WG3.ppt. Last Access: 5 September 2009.

Bétrancourt, Mireille / Tversky, Barbara. 2000. “Effect of Computer Animation on Users’ Performance: A Review.” In: Le Travail Humain. 63(4). December 2000. pp. 311–330.

Beusch, Susanna / Zäch, Olivier. 1997. “Software für die Qualitative Datenanalyse.” In: RZU Aktuell. 100. October 1997. IT Services, University of Zurich. Zurich, Switzerland. Available Online. URL: http://www.id.uzh.ch/cl/zinfo/old/editor100/editor_9.html. Last Access: 5 September 2009.

Bezjian-Avery, Alexa / Calder, Bobby / Iacobucci, Dawn. 1998. “New Media Interactive Advertising vs. Traditional Advertising.” In: Journal of Advertising Research. 38(4). July–August 1998. pp. 23–32.

Bickmore, Timothy W. 2003. "Relational Agents: Effecting Change Through Human-Computer Relationships." Ph.D. Thesis. Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. February 2003. Available Online. URL: http://affect.media.mit.edu/pdfs/03.bickmore-phd.pdf. Last Access: 5 September 2009.

Bickmore, Timothy W. 2004. "Unspoken Rules of Spoken Interaction." In: Communications of the ACM. 47(4). April 2004. pp. 38–44.

Bickmore, Timothy W. / Caruso, Lisa / Clough-Gorr, Kerri / Heeren, Tim. 2005. "'It's Just Like You Talk to a Friend': Relational Agents for Older Adults." In: Interacting with Computers. 17(6). December 2005. pp. 711–735.

Bickmore, Timothy W. / Cassell, Justine. 1999. "Small Talk and Conversational Storytelling in Embodied Conversational Interface Agents." In: Proceedings of the AAAI Fall Symposium on Narrative Intelligence. North Falmouth, Massachusetts, USA. 5–7 November 1999.

Bickmore, Timothy W. / Cassell, Justine. 2005. "Social Dialogue with Embodied Conversational Agents." In: Van Kuppevelt et al. (eds.). 2005. pp. 23–54.

Bickmore, Timothy W. / Mauer, Daniel / Crespo, Francisco / Brown, Thomas. 2008. "Negotiating Task Interruptions with Virtual Agents for Health Behavior Change." In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2008). Estoril, Portugal. 12–16 May 2008.

Bickmore, Timothy W. / Picard, Rosalind W. 2005. "Establishing and Maintaining Long-Term Human-Computer Relationships." In: ACM Transactions on Computer-Human Interaction (TOCHI). 12(2). June 2005. pp. 293–327.

Binsted, Kim. 1995. "Using Humour to Make Natural Language Interfaces More Friendly." In: Proceedings of the Workshop on Artificial Intelligence, Artificial Life, and Entertainment. 14th International Joint Conference on Artificial Intelligence (IJCAI '95). Montreal, Quebec, Canada. 20–25 August 1995.

Binsted, Kim. 1996. "Machine Humour: An Implemented Model of Puns." Ph.D. Thesis. University of Edinburgh. Edinburgh, United Kingdom. July 1996. Available Online. URL: http://www2.hawaii.edu/~binsted/papers/Binstedthesis.pdf. Last Access: 5 September 2009.

Binsted, Kim / Bergen, Benjamin / Coulson, Seana / Nijholt, Anton / Stock, Oliviero / Strapparava, Carlo / Ritchie, Graeme / Manurung, Ruli / Pain, Helen G. / Waller, Annalu / O'Mara, Dave. 2006. "Computational Humor." In: IEEE Intelligent Systems. 21(2). March/April 2006. pp. 59–69.

Binsted, Kim / Bergen, Benjamin / McKay, Justin. 2003. "Pun and Non-Pun Humor in Second-Language Learning." In: Proceedings of the Workshop on Humor Modeling in the Interface. Conference on Human Factors in Computing Systems (CHI 2003). Fort Lauderdale, Florida, USA. 5–10 April 2003.

Biocca, Frank. 1997. "The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments." In: Journal of Computer-Mediated Communication. 3(2). September 1997. Available Online. URL: http://jcmc.indiana.edu/vol3/issue2/biocca2.html. Last Access: 5 September 2009.

Biswas, Gautam / Leelawong, Krittaya / Schwartz, Daniel / Vye, Nancy / The Teachable Agents Group at Vanderbilt. 2005. "Learning by Teaching: A New Agent Paradigm for Educational Software." In: Applied Artificial Intelligence. 19(3–4). March–April 2005. pp. 363–392.

Blaikie, Norman. 2003. Analyzing Quantitative Data: From Description to Explanation. London: SAGE Publications.

Blair, David / Meyer, Tom. 1997. "Tools for an Interactive Virtual Cinema." In: Trappl & Petta. (eds.). 1997. pp. 83–91.

Blair, Kristen / Schwartz, Daniel L. / Biswas, Gautam / Leelawong, Krittaya. 2007. "Pedagogical Agents for Learning by Teaching: Teachable Agents." In: Educational Technology. 47(1). January–February 2007. pp. 56–61.

Blakey, Elaine / Spence, Sheila. 1990. "Developing Metacognition." ERIC Clearinghouse on Information Resources. Syracuse, New York, USA. ERIC Identifier ED327218. November 1990. Available Online. URL: http://www.vtaide.com/png/ERIC/Metacognition.htm. Last Access: 5 September 2009.

Blank, Moshe / Gorelick, Lena / Shechtman, Eli / Irani, Michal / Basri, Ronen. 2005. "Actions as Space-Time Shapes." In: Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV 2005). Beijing, China. 17–20 October 2005.

Blattberg, Robert C. / Deighton, John. 1991. "Interactive Marketing: Exploring the Age of Addressability." In: Sloan Management Review. 33(1). Fall 1991. pp. 5–14.

Bledsoe, Woody. 1986. "I Had a Dream: AAAI Presidential Address, 19 August 1985." In: AI Magazine. 7(1). Spring 1986. pp. 57–61.

Bloom, Benjamin S. 1984. "The 2-Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." In: Educational Researcher. 13(6). June–July 1984. pp. 4–16.

Bodomo, Adams / Luke, Kang Kwong / Anttila, Arto. 2001. "Evaluating Interactivity in Web-Based Learning." In: Proceedings of the UGC Conference on New Challenges for Educational Practice in Higher Education. Hong Kong, China. 24–26 May 2001.

Boehm, Barry W. 1988. "A Spiral Model of Software Development and Enhancement." In: IEEE Computer. 21(5). May 1988. pp. 61–72.

Boethel, Martha / Dimock, K. Victoria. 1999. "Constructing Knowledge with Technology: A Review of the Literature." Technical Report TEC-27. Southwest Educational Development Laboratory. Austin, Texas, USA. August 1999. Available Online. URL: http://www.sedl.org/pubs/tec27/constructing.pdf. Last Access: 5 September 2009.

Boettcher, Judith V. 1999. "Cyber Course Size: Pedagogy and Politics." In: Syllabus. 12(8). April 1999. pp. 42–44.

Bolt, Richard A. 1980. "'Put-that-there': Voice and Gesture at the Graphics Interface." In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques. Seattle, Washington, USA. 14–18 July 1980.

Bonato, Roberto / Bricaire, Richard / Karyda, Maria / Lanzi, Paola / Meduri, Valentino / Pandolfo, Cynthia / Riccucci, Carlo / Soria-Rodriguez, Pedro / Spanoudakis, George / Thomas, Reshma. 2007. "Evaluation Criteria." Technical Report A7.D5.1. Project SERENITY: System Engineering for Security & Dependability. 12 February 2007. Available Online. URL: http://www.serenity-forum.org/IMG/pdf/A7.D5.1_evaluation_criteria_v2.3_final.pdf. Last Access: 5 September 2009.

Bonk, Curtis J. / Dennen, Vanessa P. 2005. "Massive Multiplayer Online Gaming: A Research Framework for Military Training and Education." Technical Report 2005-1. Advanced Distributed Learning Initiative, Office of the Under Secretary of Defense for Personnel and Readiness. Washington, District of Columbia, USA. March 2005. Available Online. URL: http://blogoehlert.typepad.com/eclippings/files/GameReport_Bonk_final.pdf. Last Access: 5 September 2009.

Bonk, Curtis J. / Graham, Charles R. (eds.). 2006. The Handbook of Blended Learning: Global Perspectives, Local Designs. San Francisco: Pfeiffer.

Booch, Grady / Maksimchuk, Robert A. / Engel, Michael W. / Young, Bobbi J. / Conallen, Jim / Houston, Kelli A. 2007. Object-Oriented Analysis and Design with Applications (3rd ed.). Amsterdam: Addison-Wesley Longman.

Borland, John. 2007. "A Smarter Web: New Technologies Will Make Online Search More Intelligent – And May Even Lead to a 'Web 3.0'." In: MIT Technology Review. March–April 2007. Available Online. URL: http://www.technologyreview.com/Infotech/18306/. Last Access: 5 September 2009.

Borsook, Terry K. / Higginbotham-Wheat, Nancy. 1991. "Interactivity: What Is It and What Can It Do for Computer-Based Instruction?." In: Educational Technology. 31(10). October 1991. pp. 11–17.

Bostock, Stephen. 1995. "Classifications of Educational Software." Computer Science Department, Keele University. Keele, United Kingdom. February 1995. Available Online. URL: http://www.keele.ac.uk/depts/aa/landt/lt/docs/atcbttyp.htm. Last Access: 5 September 2009.

Bourlard, Hervé / Morgan, Nelson. 1997. "Connectionist Techniques." In: Cole et al. (eds.). 1997. pp. 356–361.

Boyer, Kristy E. / Vouk, Mladen A. / Lester, James C. 2007. "The Influence of Learner Characteristics on Task-Oriented Tutorial Dialogue." In: Proceedings of the 13th International Conference on Artificial Intelligence in Education (AI-ED 2007). Marina del Rey, California, USA. 9–13 July 2007.

Brachman, Ronald J. 1985. "'I Lied About the Trees,' Or, Defaults and Definitions in Knowledge Representation." In: AI Magazine. 6(3). Fall 1985. pp. 80–93.

Brachman, Ronald J. / Levesque, Hector J. 2004. Knowledge Representation and Reasoning. San Francisco: Morgan Kaufmann Publishers.

Bradshaw, Jeffrey M. (ed.). 1997. Software Agents. Cambridge, Massachusetts: The MIT Press.

Branham, Sheryl. 2001. "Creating Physical Personalities for Agents with Faces: Modeling Trait Impressions of the Face." In: Proceedings of the 2nd Workshop on Attitudes, Personality, and Emotions in User-Adapted Interactions. 8th International Conference on User Modeling (UM 2001). Sonthofen, Germany. 13–17 July 2001.

Branigan, Holly P. / Pickering, Martin J. / Pearson, Jamie / McLean, Janet F. / Nass, Clifford I. 2003. "Syntactic Alignment Between Computers and People: The Role of Belief About Mental States." In: Proceedings of the 25th Annual Conference of the Cognitive Science Society (CogSci 2003). Boston, Massachusetts, USA. 31 July–2 August 2003.

Bratman, Michael E. 1987. Intentions, Plans, and Practical Reason. Cambridge, Massachusetts: Harvard University Press.

Bratman, Michael E. / Israel, David J. / Pollack, Martha E. 1988. "Plans and Resource-Bounded Practical Reasoning." In: Computational Intelligence. 4(4). November 1988. pp. 349–355.

Brave, Scott / Nass, Clifford I. 2003. "Emotion in Human-Computer Interaction." In: Jacko & Sears. (eds.). 2003. pp. 81–96.

Brave, Scott / Nass, Clifford I. / Hutchinson, Kevin. 2005. "Computers that Care: Investigating the Effects of Orientation of Emotion Exhibited by an Embodied Computer Agent." In: International Journal of Human-Computer Studies. 62(2). February 2005. pp. 161–178.

Brecht, Barbara / Jones, Marlene. 1988. "Student Models: The Genetic Graph Approach." In: International Journal of Man-Machine Studies. 28(5). May 1988. pp. 483–504.

Breese, Jack / Ball, Eugene. 1998. "Modeling Emotional State and Personality for Conversational Agents." Technical Report MSR-TR-98-41. Microsoft Research. Redmond, Washington, USA. March 1998.

Brennan, Susan E. / Ohaeri, Justina O. 1994. "Effects of Message Style on Users' Attributions Toward Agents." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI '94). Boston, Massachusetts, USA. 24–28 April 1994.

Brennan, Susan E. / Schober, Michael F. 2001. "How Listeners Compensate for Disfluencies in Spontaneous Speech." In: Journal of Memory and Language. 44(2). February 2001. pp. 274–296.

Bretz, Rudy / Schmidbauer, Michael. 1983. Media for Interactive Communication. Beverly Hills, California: SAGE Publications.

Brewer, David. 1996. "A Brief Survey of Speech Recognition Technology." No Longer Available Online. URL: http://www.linfield.edu/~dbrewer/speech/spchi.html. Last Access: 12 June 1997.

Brigham, John C. 1980. "Limiting Conditions of the 'Physical Attractiveness Stereotype': Attributions About Divorce." In: Journal of Research in Personality. 14(3). September 1980. pp. 365–375.

Broadbent, Donald E. 1958. Perception and Communication. London: Pergamon Press.

Broadbent, Donald E. (ed.). 1993. The Simulation of Human Intelligence. Oxford: Blackwell Publishing.

Brooks, Jacqueline G. / Brooks, Martin G. 1993. The Case for Constructivist Classrooms. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Brooks, Rodney A. 1986. "A Robust Layered Control System for a Mobile Robot." In: IEEE Journal of Robotics and Automation. 2 (1). March 1986. pp. 14–23.

Brooks, Rodney A. 1990. "Elephants Don't Play Chess." In: Robotics and Autonomous Systems. 6(1–2). June 1990. pp. 3–15.

Brooks, Rodney A. 1991. "Intelligence Without Representation." In: Artificial Intelligence. 47(1–3). January 1991. pp. 139–159.

Brown, Andrew R. / Voltz, Bradley D. 2005. "Elements of Effective E-Learning Design." In: The International Review of Research in Open and Distance Learning. 6(1). March 2005. Available Online. URL: http://www.irrodl.org/index.php/irrodl/article/viewArticle/217/300. Last Access: 5 September 2009.

Brown, Edward / Barrett, Neil. 2006. "A Wizard-of-Oz Platform for Embodied Conversational Agents." In: Computer Animation and Virtual Worlds. 17(3–4). July 2006. pp. 249–257.

Brown, John S. / Adler, Richard P. 2008. "Minds on Fire: Open Education, the Long Tail, and Learning 2.0." In: EDUCAUSE Review. 43(1). January–February 2008. pp. 16–32. Available Online. URL: http://net.educause.edu/ir/library/pdf/ERM0811.pdf. Last Access: 5 September 2009.

Brown, John S. / Collins, Allan M. / Duguid, Paul. 1989. "Situated Cognition and the Culture of Learning." In: Educational Researcher. 18(1). January–February 1989. pp. 32–42.

Brown, John S. / VanLehn, Kurt. 1980. "Repair Theory: A Generative Theory of Bugs in Procedural Skills." In: Cognitive Science. 4(4). October 1980. pp. 379–426.

Brown, Keith. (ed.). 2006. Encyclopedia of Language and Linguistics (2^{nd} ed.). 14 Volumes. Oxford: Elsevier Science.

Brown, Penelope / Levinson, Stephen C. 1987. Politeness: Some Universals in Language Usage. Cambridge: Cambridge University Press.

Brown, Terry D. / Dane, Francis C. / Durham, Marcus D. 1998. "Perception of Race and Ethnicity." In: Journal of Social Behavior and Personality. 13(2). 1998. pp. 295–306.

Bruner, Jerome S. 1961. "The Act of Discovery." In: Harvard Educational Review. 31(1). 1961. pp. 21–32.

Bruner, Jerome S. 1986. Actual Minds, Possible Worlds. Cambridge, Massachusetts: Harvard University Press.

Bruner, Jerome S. 1990. Acts of Meaning: Four Lectures on Mind and Culture. Cambridge, Massachusetts: Harvard University Press.

Bruner, Jerome S. 1991. "The Narrative Construction of Reality." In: Critical Inquiry. 18(1). Autumn 1991. pp. 1–21.

Brusilovsky, Peter. 1992. "A Framework for Intelligent Knowledge Sequencing and Task Sequencing." In: Proceedings of the 2nd International Conference on Intelligent Tutoring Systems (ITS '92). Montreal, Quebec, Canada. 10–12 June 1992.

Brusilovsky, Peter. 1994. "The Construction and Application of Student Models in Intelligent Tutoring Systems." In: Journal of Computer and Systems Sciences International. 32(1). 1994. pp. 70–89.

Bucy, Erik P. 2004a. "The Debate." In: The Information Society. 20(5). November–December 2004. p. 371.

Bucy, Erik P. 2004b. "Interactivity in Society: Locating an Elusive Concept." In: The Information Society. 20(5). November–December 2004. pp. 373–383.

Buisine, Stéphanie / Abrilian, Sarkis / Martin, Jean-Claude. 2004. "Evaluation of Multimodal Behaviour of Embodied Agents: Cooperation Between Speech and Gestures." In: Ruttkay & Pelachaud. (eds.). 2004. pp. 217–238.

Buisine, Stéphanie / Abrilian, Sarkis / Rendu, Christophe / Martin, Jean-Claude. 2002. "Towards Experimental Specification and Evaluation of Lifelike Multimodal Behavior." In: Proceedings of the Workshop "Embodied Conversational Agents – Let's Specify and Evaluate Them!." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Buisine, Stéphanie / Martin, Jean-Claude. 2003. "Experimental Evaluation of Bi-Directional Multimodal Interaction with Conversational Agents." In: Proceedings of the 9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003). Zurich, Switzerland. 1–5 September 2003.

Buisine, Stéphanie / Martin, Jean-Claude. 2005. "Children's and Adults' Multimodal Interaction with 2D Conversational Agents." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2005). Portland, Oregon, USA. 2–7 April 2005.

Bull, Susan. 1998. "'Do It Yourself' Student Models for Collaborative Student Modeling and Peer Interaction." In: Proceedings of the 4th International Conference on Intelligent Tutoring Systems (ITS '98). San Antonio, Texas, USA. 16–19 August 1998.

Bull, Susan / Abu-Issa, Abdallatif S. / Ghag, Harpreet / Lloyd, Tim. 2005. "Some Unusual Open Learner Models." In: Proceedings of the 12th International Conference on Artificial Intelligence in Education (AI-ED 2005). Amsterdam, The Netherlands. 18–22 July 2005.

Bull, Susan / Brna, Paul / Critchley, Sonia / Davie, Koula / Holzherr, Corina. 1999. "The Missing Peer, Artificial Peers and the Enhancement of Human-Human Collaborative Student Modelling." In: Proceedings of the 9th International Conference on Artificial Intelligence in Education (AI-ED '99). LeMans, France. 19–23 July 1999.

Bull, Susan / Dimitrova, Vania / McCalla, Gordon I. (eds.). 2007a. Open Learner Models. International Journal of Artificial Intelligence in Education. Special Double Issue. 17(2–3). 2007.

Bull, Susan / Dimitrova, Vania / McCalla, Gordon I. 2007b. "Preface for 17(2)." In: International Journal of Artificial Intelligence in Education. 17(2). 2007. pp. 83–87.

Bull, Susan / Pain, Helen G. 1995. "Did I Say What I Think I Said, and Do You Agree with Me?: Inspecting and Questioning the Student Model." In: Proceedings of the 7th World Conference on Artificial Intelligence in Education (AI-ED '95). Washington, District of Columbia, USA. 16–19 August 1995.

Burdea, Gregoire C. / Coiffet, Philippe. 2003. Virtual Reality Technology (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Burgoon, Judee K. / Bonito, Joseph A. / Bengtsson, Björn / Cederberg, Carl / Lundeberg, Magnus / Allspach, Lisa. 2000. "Interactivity in Human-Computer Interaction: A Study of Credibility, Understanding, and Influence." In: Computers in Human Behavior. 16(6). November 2000. pp. 553–574.

Burgoon, Judee K. / Bonito, Joseph A. / Ramirez, Artemio, Jr. / Dunbar, Norah E. / Kam, Karadeen / Fischer, Jenna. 2002. "Testing the Interactivity Principle: Effects of Mediation, Propinquity, and Verbal and Nonverbal Modalities in Interpersonal Interaction." In: Journal of Communication. 52(3). September 2002. pp. 657–677.

Burleson, Winslow. 2005. "Developing Creativity, Motivation, and Self-Actualization with Learning Systems." In: International Journal of Human-Computer Studies. 63(4–5). October 2005. pp. 436–451.

Burleson, Winslow / Picard, Rosalind W. 2004. "Affective Agents: Sustaining Motivation to Learn Through Failure and a State of 'Stuck'." In: Proceedings of the Workshop on Social and Emotional Intelligence in Learning Environments. 7th International Conference on Intelligent Tutoring Systems (ITS 2004). Maceió, Alagoas, Brazil. 30 August–3 September 2004.

Burnham, Byron R. / Walden, Beth. 1997. "Interactions in Distance Education: A Report from the Other Side." In: Proceedings of the Adult Education Research Conference (AERC '97). Stillwater, Oklahoma, USA. 16–18 May 1997.

Burton, Richard R. 1982. "Diagnosing Bugs in a Simple Procedural Skill." In: Sleeman & Brown. (eds.). 1982. pp. 157–183.

Burton, Richard R. / Brown, John S. 1977. "Semantic Grammar: A Technique for Constructing Natural Language Interfaces to Instructional Systems." BBN Report 3587. Bolt Beranek and Newman, Inc. Cambridge, Massachusetts, USA. May 1977. Available Online. URL: http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED142240. Last Access: 5 September 2009.

Burton, Richard R. / Brown, John S. 1982. "An Investigation of Computer Coaching for Informal Learning Activities." In: Sleeman & Brown. (eds.). 1982. pp. 79–98.

Cahn, Janet E. 1990. "The Generation of Affect in Synthesized Speech." In: Journal of the American Voice I/O Society. 8. July 1990. pp. 1–19.

Cai, Yang / Abascal, Julio. (eds.). 2006. Ambient Intelligence in Everyday Life. Berlin: Springer Verlag.

Campbell, Donald T. / Stanley, Julian C. 1963. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.

Cañamero, Lola D. 2003. "Designing Emotions for Activity Selection in Autonomous Agents." In: Trappl et al. (eds.). 2003. pp. 115–148.

Cañamero, Lola D. / Aylett, Ruth S. (eds.). 2008. Animating Expressive Characters for Social Interaction. Amsterdam: John Benjamins.

Cann, Arnie / Calhoun, Lawrence G. / Banks, Janet S. 1997. "On the Role of Humor Appreciation in Interpersonal Attraction: It's No Joking Matter." In: Humor: International Journal of Humor Research. 10(1). 1997. pp. 77–89.

Carberry, Sandra. 2001. "Techniques for Plan Recognition." In: User Modeling and User-Adapted Interaction. 11(1–2). March 2001. pp. 31–48.

Carbonell, Jaime R. 1970. "AI in CAI: An Artificial Intelligence Approach to Computer-Assisted Instruction." In: IEEE Transactions on Man-Machine Systems. 11(4). December 1970. pp. 190–202.

Carey, John. 1989. "Interactive Media." In: Barnouw. (ed.). 1989. Volume 2. pp. 328–330.

Carr, Brian / Goldstein, Ira P. 1977. "Overlays: A Theory of Modeling for Computer-Aided Instruction." AI Memo 406. Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. February 1977. Available Online. URL: ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-406.pdf. Last Access: 5 September 2009.

Carroll, John M. (ed.). 2001. Human-Computer Interaction in the New Millennium. Amsterdam: Addison-Wesley Longman.

Carroll, John. 2003. "Parsing." In: Mitkov. (ed.). 2003. pp. 233–248.

Cassell, Justine. 2000a. "Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents." In: Cassell et al. (eds.). 2000a. pp. 1–27.

Cassell, Justine. 2000b. "Embodied Conversational Interface Agents." In: Communications of the ACM. 43(4). April 2000. pp. 70–78.

Cassell, Justine. 2001. "Embodied Conversational Agents: Representation and Intelligence in User Interfaces." In: AI Magazine. 22(4). Winter 2001. pp. 67–84.

Cassell, Justine. 2007. "Body Language: Lessons from the Near-Human." In: Riskin. (ed.). 2007. pp. 346–374.

Cassell, Justine / Bickmore, Timothy W. 2000. "External Manifestations of Trustworthiness in the Interface." In: Communications of the ACM. 43(12). December 2000. pp. 50–56.

Cassell, Justine / Bickmore, Timothy W. / Campbell, Lee / Chang, Kenny / Vilhjálmsson, Hannes H. / Yan, Hao. 1999. "Requirements for an Architecture for Embodied Conversational Characters." In: Proceedings of the 10th Eurographics Workshop on Animation and Simulation. Milan, Italy. 7–8 September 1999.

Cassell, Justine / Nakano, Yukiko I. / Bickmore, Timothy W. / Sidner, Candace L. / Rich, Charles. 2001a. "Annotating and Generating Posture from Discourse Structure in Embodied Conversational Agents." In: Proceedings of the Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents. 5th International Conference on Autonomous Agents (Agents 2001). Montreal, Quebec, Canada. 28 May–01 June 2001.

Cassell, Justine / Vilhjálmsson, Hannes H. / Bickmore, Timothy W. 2001b. "BEAT: The Behavior Expression Animation Toolkit." In: Proceedings of the 28th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001). Los Angeles, California, USA. 12–17 August 2001.

Cassell, Justine / Pelachaud, Catherine / Badler, Norman I. / Steedman, Mark / Achorn, Brett / Becket, Tripp / Douville, Brett / Prevost, Scott / Stone, Matthew. 1994. "Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents." In: Proceedings of the 21st International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1994). Orlando, Florida, USA. 24–29 July 1994.

Cassell, Justine / Stone, Matthew D. 1999. "Living Hand to Mouth: Psychological Theories About Speech and Gesture in Interactive Dialogue Systems." In: Proceedings of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems. Cape Cod, Massachusetts, USA. 5–7 November 1999.

Cassell, Justine / Sullivan, Joseph W. / Prevost, Scott / Churchill, Elizabeth F. (eds.). 2000a. Embodied Conversational Agents. Cambridge, Massachusetts: The MIT Press.

Cassell, Justine / Bickmore, Timothy W. / Campbell, Lee / Vilhjálmsson, Hannes H. / Yan, Hao. 2000b. "Human Conversation as a System Framework: Designing Embodied Conversational Agents." In: Cassell et al. (eds.). 2000a. pp. 29–63.

Cassell, Justine / Tartaro, Andrea. 2007. "Intersubjectivity in Human-Agent Interaction." In: Interaction Studies. 8(3). 2007. pp. 391–410.

Cassell, Justine / Tartaro, Andrea / Rankin, Yolanda / Oza, Vani / Tse, Candice. 2007. "Virtual Peers for Literacy Learning." In: Educational Technology. 47(1). January–February 2007. pp. 39–42.

Cassell, Justine / Thórisson, Kristinn R. 1999. "The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents." In: Applied Artificial Intelligence. 13(4–5). May 1999. pp. 519–538.

Castronova, Edward. 2005, 2006. Synthetic Worlds: The Business and Culture of Online Games. Paperback Edition. Chicago: The University of Chicago Press.

Catrambone, Richard / Stasko, John / Xiao, Jun. 2004. "ECA as User Interface Paradigm: Experimental Findings within a Framework for Research." In: Ruttkay & Pelachaud. (eds.). 2004. pp. 239–267.

Cawsey, Alison. 1998. The Essence of Artificial Intelligence. Harlow: Prentice Hall.

Chaffey, Dave. 2006. E-Business and E-Commerce Management (3rd ed.). New York: Financial Times / Prentice Hall.

Chandler, Daniel. 2002. Semiotics: The Basics. London: Routledge.

Chang, Shi-Kuo. (ed.). 2001. Handbook of Software Engineering and Knowledge Engineering. Volume 1. Singapore: World Scientific Publishing.

Chapman, Alan. 2002. "Hierarchy of Needs (1990's Eight-Stage Model Based on Maslow)." Alan Chapman Consultancy. Leicester, United Kingdom. Available Online. URL: http://www.businessballs.com/maslowhierarchyofneeds8.pdf. Last Access: 5 September 2009.

Charmaz, Kathy C. 2006. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Thousand Oaks, California: SAGE Publications.

Charniak, Eugene. 1991. "Bayesian Networks Without Tears." In: AI Magazine. 12(4). Winter 1991. pp. 50–63.

Charniak, Eugene. 1996. Statistical Language Learning. Cambridge, Massachusetts: The MIT Press.

Charniak, Eugene. 1997. "Statistical Techniques for Natural Language Parsing." In: AI Magazine. 18(4). Winter 1997. pp. 33–43.

Charniak, Eugene. 2001. "Statistical Techniques in Natural Language Processing." In: Wilson & Keil. (eds.). 2001. pp. 801–802.

Chateau, Noël / Maffiolo, Valérie / Pican, Nathalie / Mersiol, Marc. 2005. "The Effect of Embodied Conversational Agents' Speech Quality on Users' Attention and Emotion." In: Proceedings of the

1st International Conference on Affective Computing and Intelligent Interaction (ACII 2005). Beijing, China. 22–24 October 2005.

Chen, Guanling / Kotz, David. 2000. “A Survey of Context-Aware Mobile Computing Research.” Dartmouth Computer Science Technical Report TR2000-381. Department of Computer Science, Dartmouth College. Hanover, New Hampshire, USA. November 2000. Available Online. URL: http://www.cs.dartmouth.edu/reports/TR2000-381.pdf. Last Access: 6 September 2009.

Chen, Peter Y. / Krauss, Autumn D. 2004. “Simple Correlation (Regression).” In: Lewis-Beck et al. (eds.). 2004. p. 1035.

Chittaro, Luca. 2006. “Visualizing Information on Mobile Devices.” In: Computer. 39(3). March 2006. pp. 40–45.

Chomsky, Noam. 1956. “Three Models for the Description of Language.” In: IRE Transactions on Information Theory. 2(3). September 1956. pp. 113–124.

Chomsky, Noam. 1957. Syntactic Structures. The Hague: Mouton.

Chomsky, Noam. 1959. “A Review of B. F. Skinner’s Verbal Behavior.” In: Language. 35(1). 1959. pp. 26–58.

Chou, Chien. 2003. “Interactivity and Interactive Functions in Web-Based Learning Systems: A Technical Framework for Designers.” In: British Journal of Educational Technology. 34(3). June 2003. pp. 265–279.

Chou, Chih-Yueh / Chan, Tak-Wai / Lin, Chi-Jen. 2003. “Redefining the Learning Companion: The Past, Present, and Future of Educational Agents.” In: Computers & Education. 40(3). April 2003. pp. 255–269.

Chrisley, Ronald L. 2004. “Artificial Intelligence.” In: Gregory. (ed.). 2004. pp. 61–63.

Christensen, Larry B. 2006. Experimental Methodology (10th ed.). Boston, Massachusetts: Allyn & Bacon.

Christoph, Noor. 2004. “Empirical Evaluation Methodology for Embodied Conversational Agents: On Conducting Evaluation Studies.” In: Ruttkay & Pelachaud. (eds.). 2004. pp. 67–99.

Churchill, Elizabeth F. / Cook, Linda / Hodgson, Peter / Prevost, Scott / Sullivan, Joseph W. 2000. “‘May I Help You?’: Designing Embodied Conversational Agent Allies.” In: Cassell et al. (eds.). 2000a. pp. 64–94.

Clancey, William J. 1986. “Qualitative Student Models.” In: Annual Review of Computer Science. 1. June 1986. pp. 381–450.

Clarebout, Geraldine / Elen, Jan / Johnson, W. Lewis / Shaw, Erin. 2002. “Animated Pedagogical Agents: An Opportunity to be Grasped?.” In: Journal of Educational Multimedia and Hypermedia. 11(3). September 2002. pp. 267–286.

Clark, Andy / Sloman, Aaron / Zeman, Adam / Oberlander, Jon. 2004a. “Can Computers Have Emotions?: The Debate Panel – Position Statements.” 17 November 2004. Available Online. URL: http://www.inf.ed.ac.uk/events/hotseat/panel_statements.html. Last Access: 6 September 2009.

Clark, Kendall / Parsia, Bijan / Hendler, James A. 2004b. “Will the Semantic Web Change Education?.” In: Journal of Interactive Media in Education. Special Issue on the Educational Semantic Web. 2004(3). 21 May 2004. Available Online. URL: http://www-jime.open.ac.uk/2004/3/clark-2004-3.pdf. Last Access: 6 September 2009.

Clark, Ruth C. 2002. “Six Principles of Effective E-Learning: What Works and Why.” In: The E-Learning Developers’ Journal. 10 September 2002. pp. 1–8.

Clark, Ruth C. / Lyons, Chopeta. 2004. Graphics for Learning: Proven Guidelines for Planning, Designing, and Evaluating Visuals in Training Materials. San Francisco: Pfeiffer.

Clark, Ruth C. / Mayer, Richard E. 2003. E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. San Francisco: Pfeiffer.

Clark, Ruth C. / Nguyen, Frank / Sweller, John. 2006. Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load. San Francisco: Pfeiffer.

Cliff, Dave / Grand, Stephen. 1999. “The Creatures Global Digital Ecosystem.” In: Artificial Life. 5(1). Winter 1999. pp. 77–93.

Clouse, R. Wilburn / Spurgeon, Karen L. 1995. “Corporate Analysis of Humor.” In: Psychology: A Journal of Human Behavior. 32(3–4). 1995. pp. 1–24.

Coan, James A. / Allen, John J. B. (eds.). 2007. Handbook of Emotion Elicitation and Assessment. New York: Oxford University Press.

Cohen, Philip R. 1997. “Dialogue Modeling.” In: Cole et al. (eds.). 1997. pp. 204–210.

Cohen, Philip R. / Perrault, C. Raymond / Allen, James F. 1982. “Beyond Question Answering.” In: Lehnert & Ringle. (eds.). 1982. pp. 245–274.

Cohn, Jeffrey F. / Kanade, Takeo. 2007. “Use of Automated Facial Image Analysis for Measurement of Emotion Expression.” In: Coan & Allen. (eds.). 2007. pp. 222–238.

Colburn, R. Alex / Cohen, Michael F. / Drucker, Steven M. 2000. “The Role of Eye Gaze in Avatar Mediated Conversational Interfaces.” Technical Report MSR-TR-2000-81. Microsoft Research. Redmond, Washington, USA. 31 July 2000. Available Online. URL: ftp://ftp.research.microsoft.com/pub/tr/tr-2000-81.pdf. Last Access: 6 September 2009.

Cole, Peter / Morgan, Jerry L. (eds.). 1975. Syntax and Semantics Volume 3: Speech Acts. New York: Academic Press.

Cole, Ronald A. / Hirschman, Lynette (eds.). / Atlas, Les / Beckman, Mary / Biermann, Alan / Bush, Marcia / Clements, Mark / Cohen, Jordan / Garcia, Oskar / Hanson, Brian / Hermansky, Hynek / Levinson, Steve / McKeown, Kathleen R. / Morgan, Nelson / Novick, David G. / Ostendorf, Mari / Oviatt, Sharon / Price, Patti / Silverman, Harvey / Spitz, Judy / Waibel, Alexander / Weinstein, Clifford / Zahorian, Steve / Zue, Victor W. 1995. “The Challenge of Spoken Language Systems: Research Directions for the Nineties.” In: IEEE Transactions on Speech and Audio Processing. 3(1). January 1995. pp. 1–20.

Cole, Ronald A. / Mariani, Joseph / Uszkoreit, Hans / Varile, Giovanni B. / Zaenen, Annie / Zampolli, Antonio / Zue, Victor W. (eds.). 1997. Survey of the State of the Art in Human Language Technology. Cambridge: Cambridge University Press.

Cole, Ronald A. / van Vuuren, Sarel / Pellom, Bryan / Hacioglu, Kadri / Ma, Jiyong / Movellan, Javier / Schwartz, Scott / Wade-Stein, David / Ward, Wayne / Yan, Jie. 2003. “Perceptive Animated Interfaces: First Steps Toward a New Paradigm for Human-Computer Interaction.” In: Proceedings of the IEEE. 91(9). September 2003. pp. 1391–1405.

Coleman, John. 2005. Introducing Speech and Language Processing. Cambridge: Cambridge University Press.

Collins, Allan M. / Brown, John S. / Holum, Ann. 1991. “Cognitive Apprenticeship: Making Thinking Visible.” In: American Educator. 15(3). Winter 1991. pp. 6–11 and 38–46.

Collins, Allan M. / Quillian, M. Ross. 1969. “Retrieval Time from Semantic Memory.” In: Journal of Verbal Learning and Verbal Behavior. 8(2). April 1969. pp. 240–247.

Conati, Cristina / Gertner, Abigail S. / VanLehn, Kurt. 2002. “Using Bayesian Networks to Manage Uncertainty in Student Modeling.” In: User Modeling and User-Adapted Interaction. 12(4). November 2002. pp. 371–417.

Conati, Cristina / Gertner, Abigail S. / VanLehn, Kurt / Druzdzel, Marek J. 1997. “On-Line Student Modeling for Coached Problem Solving Using Bayesian Networks.” In: Proceedings of the 6th International Conference on User Modeling (UM ’97). Chia Laguna, Sardinia, Italy. 2–5 June 1997.

Coniam, David. 2008. “Evaluating the Language Resources of Chatbots for Their Potential in English as a Second Language.” In: ReCALL. 20(1). January 2008. pp. 98–116.

Coolidge, Frederick L. 2006. Statistics: A Gentle Introduction (2nd ed.). Thousand Oaks, California: SAGE Publications.

Cooper, Jean H. 2005. “A Training Programme Based on the Principles of Social Constructivism and Focused on Developing People for the Future World of Work: An Evaluation.” Magisters Commercii Thesis. Faculty of Economic and Business Sciences, University of Pretoria. Pretoria, South Africa. January 2005. Available Online. URL: http://upetd.up.ac.za/thesis/available/etd-03162005-142319/ unrestricted/00dissertation.pdf. Last Access: 6 September 2009.

Copeland, B. Jack. 1993. Artificial Intelligence: A Philosophical Introduction. Oxford: Blackwell Publishing.

Copeland, B. Jack. 2000. “The Turing Test.” In: Minds and Machines. 10(4). November 2000. pp. 519–539.

Corbett, Albert T. / Anderson, John R. 1994. “Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge.” In: User Modeling and User-Adapted Interaction. 4(4). December 1994. pp. 253–278.

Corbett, Albert T. / Bhatnagar, Akshat. 1997. “Student Modeling in the ACT Programming Tutor: Adjusting a Procedural Learning Model With Declarative Knowledge.” In: Proceedings of the 6th International Conference on User Modeling (UM ’97). Chia Laguna, Sardinia, Italy. 2–5 June 1997.

Corbett, Albert T. / Koedinger, Kenneth R. / Anderson, John R. 1997. “Intelligent Tutoring Systems.” In: Helander et al. (eds.). 1997. pp. 849–874.

Corkill, Daniel D. 1991. “Blackboard Systems.” In: AI Expert. 6(9). September 1991. pp. 40–47.

Corkill, Daniel D. 2003. “Collaborating Software: Blackboard and Multi-Agent Systems and the Future.” In: Proceedings of the International Lisp Conference (ILC 2003). New York City, New York, USA. 12–15 October 2003.

Costa, Ernesto. (ed.). 1992. New Directions for Intelligent Tutoring Systems. Berlin: Springer Verlag.

Costa, Paul T. / McCrae, Robert R. 1992. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual. Odessa, Florida: Psychological Assessment Resources.

Costabile, Maria F. 2001. “Usability in the Software Life Cycle.” In: Chang. (ed.). 2001. pp. 179–192.

Cowell, Andrew J. / Stanney, Kay M. 2003. “Embodiment and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents.” In: Proceedings of the 4th International Working Conference on Intelligent Virtual Agents (IVA 2003). Kloster Irsee, Germany. 15–17 September 2003.

Cowie, Jim / Wilks, Yorick. 2000. “Information Extraction.” In: Dale et al. (eds.). 2000. pp. 241–260.

Craig, John J. 2004. Introduction to Robotics: Mechanics and Control (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Craig, Scotty D. / Gholson, Barry / Driscoll, David M. 2002. “Animated Pedagogical Agents in Multimedia Educational Environments: Effects of Agent Properties, Picture Features, and Redundancy.” In: Journal of Educational Psychology. 94(2). June 2002. pp. 428–434.

Craig, Scotty D. / Graesser, Arthur C. / Sullins, Jeremiah / Gholson, Barry. 2004. “Affect and Learning: An Exploratory Look into the Role of Affect in Learning with AutoTutor.” In: Journal of Educational Media. 29(3). October 2004. pp. 241–250.

Von Cranach, Mario / Foppa, Klaus / Lepenies, Wolf / Ploog, Detlev. (eds.). 1979. Human Ethology: Claims and Limits of a New Discipline. Cambridge: Cambridge University Press.

Crossley, Michele L. 2000. Introducing Narrative Psychology: Self, Trauma, and the Construction of Meaning. Buckingham, United Kingdom: Open University Press.

Crowley, Terry. 2007. Field Linguistics: A Beginner’s Guide. New York: Oxford University Press.

Crowther, Paul / Hartnett, Jackie / Williams, Ray. 1995. “Teaching Repertory Grid Concepts for Knowledge Acquisition in Expert Systems: An Interactive Approach.” In: Proceedings of the 12th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE ’95). Melbourne, Australia. 4–6 December 1995.

Cruse, D. Alan. 2000. Meaning in Language: An Introduction to Semantics and Pragmatics. Oxford: Oxford University Press.

Csikszentmihalyi, Mihaly. 1975. Beyond Boredom and Anxiety: Experiencing Flow in Work and Play. San Francisco: Jossey-Bass.

Csikszentmihalyi, Mihaly. 1990. Flow: The Psychology of Optimal Experience. New York: Harper & Row.

Cullingford, Richard E. 1978. “Script Application: Computer Understanding of Newspaper Stories.” Ph.D. Thesis. Department of Computer Science, Yale University. New Haven, Connecticut, USA. Research Report No. 116. January 1978.

Cunningham, Donald J. 1991. “Assessing Constructions and Constructing Assessments: A Dialogue.” In: Educational Technology. 31(5). May 1991. pp. 13–17.

Cunningham, Hamish. 2006. “Information Extraction, Automatic.” In: Brown. (ed.). 2006. Volume 5. pp. 665–677.

D’Alessandro, Christophe / Liénard, Jean-Sylvain. 1997. “Synthetic Speech Generation.” In: Cole et al. (eds.). 1997. pp. 170–175.

D’Ambrosio, Bruce. 1999. “Inference in Bayesian Networks.” In: AI Magazine. 20(2). Summer 1999. pp. 21–36.

D’Mello, Sidney K. / Craig, Scotty D. / Sullins, Jeremiah / Graesser, Arthur C. 2006. “Predicting Affective States Expressed Through an Emote-Aloud Procedure from AutoTutor’s Mixed-Initiative Dialogue.” In: International Journal of Artificial Intelligence in Education. 16(1). 2006. pp. 3–28.

D’Mello, Sidney K. / Craig, Scotty D. / Witherspoon, Amy / McDaniel, Bethany / Graesser, Arthur C. 2008. “Automatic Detection of Learner’s Affect from Conversational Cues.” In: User Modeling and User-Adapted Interaction. 18(1–2). February 2008. pp. 45–80.

D'Mello, Sidney K. / Graesser, Arthur C. 2007. "Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments." In: Proceedings of the 13th International Conference on Artificial Intelligence in Education (AI-ED 2007). Marina del Rey, California, USA. 9–13 July 2007.

D'Mello, Sidney K. / Picard, Rosalind W. / Graesser, Arthur C. 2007a. "Towards an Affect-Sensitive AutoTutor." In: IEEE Intelligent Systems. 22(4). July 2007. pp. 53–61.

D'Mello, Sidney K. / Taylor, Roger S. / Graesser, Arthur C. 2007b. "Monitoring Affective Trajectories During Complex Learning." In: Proceedings of the 29th Annual Conference of the Cognitive Science Society (CogSci 2007). Nashville, Tennessee, USA. 1–4 August 2007.

Dahlbäck, Nils / Jönsson, Arne / Ahrenberg, Lars. 1993. "Wizard of Oz Studies – Why and How." In: Knowledge-Based Systems. 6(4). December 1993. pp. 258–266.

Dale, Robert / Moisl, Hermann / Somers, Harold. (eds.). 2000. Handbook of Natural Language Processing. New York: Marcel Dekker.

Dalgarno, Barney. 1996. "Constructivist Computer Assisted Learning: Theory and Techniques." In: Proceedings of the 13th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE '96). Adelaide, Australia. 2–4 December 1996.

Dalgleish, Tim / Power, Mick. (eds.). 1999. Handbook of Cognition and Emotion. Chichester: John Wiley & Sons.

Damasio, Antonio R. 1994. Descartes' Error: Emotion, Reason, and the Human Brain. New York: Avon Books.

Daniel, John S. / Marquis, Clement. 1988. "Interaction and Independence: Getting the Mix Right." In: Sewart et al. (eds.). 1988. pp. 339–359.

Dara-Abrams, Benay P. 2002. "Theoretical Foundation in Educational Psychology for Multi-Intelligent Online Learning." BrainJolt. Los Altos, California, USA. June 2002. Available Online. URL: http://www.brainjolt.com/docs/foundation.pdf. Last Access: 6 September 2009.

Darby, John K. (ed.). 1981. Speech Evaluation in Psychiatry. New York: Grune & Stratton.

Darwin, Charles. 1872. The Expression of the Emotions in Man and Animals. London: John Murray.

Das, Dipanjan / Martins, André F. T. 2007. "A Survey on Automatic Text Summarization." Literature Survey for the Course Language and Statistics II (11-762), Fall 2007. Language Technologies Institute, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. 21 November 2007. Available Online. URL: http://www.cs.cmu.edu/~dipanjan/pubs/summarization.pdf. Last Access: 6 September 2009.

Dash, Anil. 2002. "Introducing the Microcontent Client." 13 November 2002. No Longer Available Online. URL: http://www.anildash.com/magazine/2002/11/introducing_the.html. Last Access: 22 November 2008.

Dautenhahn, Kerstin. 1999. "Socially Situated Life-Like Agents: If It Makes You Happy Then It Can't Be That Bad?!." In: Proceedings of the Virtual Worlds and Simulation Conference (VWSIM '99). Part of the SCS Western Multiconference. San Francisco, California, USA. 17–20 January 1999.

Dautenhahn, Kerstin. (ed.). 2000. Human Cognition and Social Agent Technology. Amsterdam: John Benjamins.

Dautenhahn, Kerstin. 2004. "Socially Intelligent Agents in Human Primate Culture." In: Payr & Trappl. (eds.). 2004. pp. 45–71.

Davenport, Thomas H. / De Long, David W. / Beers, Michael C. 1998. "Successful Knowledge Management Projects." In: Sloan Management Review. 39(2). Winter 1998. pp. 43–57.

Davidson, Richard J. / Scherer, Klaus R. / Goldsmith, H. Hill. (eds.). 2003. Handbook of the Affective Sciences. New York: Oxford University Press.

Dean, Peter J. / Stahl, Michael J. / Sylwester, David L. / Peat, Jillian A. 2001. “Effectiveness of Combined Delivery Modalities for Distance Learning and Resident Learning.” In: Quarterly Review of Distance Education. 2(3). Fall 2001. pp. 247–254.

De Angeli, Antonella. 2005. “To the Rescue of a Lost Identity: Social Perception in Human-Chatterbot Interaction.” In: Proceedings of the Joint Symposium on Virtual Social Agents. Artificial Intelligence and Simulation of Behavior (AISB) Annual Convention on Social Intelligence and Interaction in Animals, Robots, and Agents (AISB 2005). 12–15April 2005. Hatfield, United Kingdom.

De Angeli, Antonella / Johnson, Graham I. / Coventry, Lynn. 2001a. “The Unfriendly User: Exploring Social Reactions to Chatterbots.” In: Proceedings of the International Conference on Affective Human Factors Design (CAHD 2001). Singapore. 27–29 June 2001.

De Angeli, Antonella / Lynch, Paula / Johnson, Graham I. 2001b. “Personifying the E-Market: A Framework for Social Agents.” In: Proceedings of the 8th TC13 IFIP International Conference on Human-Computer Interaction (INTERACT 2001). Tokyo, Japan. 9–13 July 2001.

Dede, Christopher. 2004. “Planning for ‘Neomillennial’ Learning Styles: Implications for Investments in Technology and Faculty.” Unpublished Paper. Harvard Graduate School of Education. Harvard University. Cambridge, Massachusetts, USA. August 2004. Available Online. URL: http://www.gse.harvard.edu/~dedech/DedeNeoMillennial.pdf. Last Access: 6 September 2009.

Dede, Christopher / Lewis, Matthew. 1995. “Assessment of Emerging Educational Technologies that Might Assist and Enhance School-to-Work Transitions.” National Technical Information Service. Washington, District of Columbia, USA. Available Online. URL: http://www.virtual.gmu.edu/pdf/ota.pdf. Last Access: 6 September 2009.

Deely, John. 2005. Basics of Semiotics (4th ed.). Tartu: Tartu University Press.

DeFleur, Melvin L. / Ball-Rokeach, Sandra J. 1989. Theories of Mass Communication (5th ed.). New York: Longman.

Dehn, Doris M. / van Mulken, Susanne. 2000. “The Impact of Animated Interface Agents: A Review of Empirical Research.” In: International Journal of Human-Computer Studies. 52(1). January 2000. pp. 1–22.

Deighton, John. 1996. “The Future of Interactive Marketing.” In: Harvard Business Review. 74(6). November–December 1996. pp. 131–161.

De Jong, Ton / Sarti, Luigi. (eds.). 1994. Design and Production of Multimedia and Simulation-Based Learning Material. Dordrecht, The Netherlands: Kluwer Academic Publishers.

DeKanter, Nick. 2005. “Gaming Redefines Interactivity for Learning.” In: TechTrends: Linking Research & Practice to Improve Learning. 49(3). May–June 2005. pp. 26–32.

Delac, Kresimir / Grgic, Mislav. (eds.). 2007. Face Recognition. Vienna: I-Tech Education and Publishing.

Dempster, Arthur P. 1968. “A Generalization of Bayesian Inference.” In: Journal of the Royal Statistical Society. Series B. 30(2). 1968. pp. 205–247.

Denis, Brigitte / Watland, Philip / Pirotte, Sébastien / Verday, Nathalie. 2004. “Roles and Competencies of the e-Tutor.” In: Proceedings of the Networked Learning Conference (NLC 2004). Lancaster, United Kingdom. 5–7 April 2004.

Dennett, Daniel C. 1987. The Intentional Stance. Cambridge, Massachusetts: The MIT Press.

Dennett, Daniel C. 2001. “Intentional Stance.” In: Wilson & Keil. (eds.). 2001. pp. 412–413.

Denzin, Norman K. / Lincoln, Yvonna S. 2003. The Landscape of Qualitative Research: Theories and Issues (2nd ed.). Thousand Oaks, California: SAGE Publications.

De Rosis, Fiorella / Pelachaud, Catherine / Poggi, Isabella. 2004. “Transcultural Believability in Embodied Agents: A Matter of Consistent Adaptation.” In: Payr & Trappl. (eds.). 2004. pp. 75–105.

Devedzic, Vladan. 2004. “Education and the Semantic Web.” In: International Journal of Artificial Intelligence in Education. 14(2). 2004. pp. 165–191.

DeWalt, Kathleen M. / DeWalt, Billie R. 2002. Participant Observation: A Guide for Fieldworkers. Walnut Creek, California: AltaMira Press.

Dewey, John. 1916. Democracy and Education. New York: MacMillan Publishers. Available Online. URL: http://www.ilt.columbia.edu/publications/dewey.html. Last Access: 6 September 2009.

Dickinson, David. 1995. “Multimedia Myths.” In: Australian Personal Computer. 16(10). 1995. pp. 144–145.

Dietterich, Thomas G. 2003. “Machine Learning.” Department of Computer Science, Oregon State University. Corvallis, Oregon, USA. Available Online. URL: http://web.engr.oregonstate.edu/~tgd/publications/nature-ecs-machine-learning.pdf. Last Access: 6 September 2009.

Dietterich, Thomas G. / Kong, Eun B. 1995. “Machine Learning Bias, Statistical Bias, and Statistical Variance of Decision Tree Algorithms.” Technical Report. Department of Computer Science, Oregon State University. Corvallis, Oregon, USA. Available Online. URL: http://www.cs.orst.edu/~tgd/publications/tr-bias.ps.gz. Last Access: 6 September 2009.

Digman, John M. 1990. “Personality Structure: Emergence of the Five-Factor Model.” In: Annual Review of Psychology. 41. January 1990. pp. 417–440.

Dillman, Don A. 2007. Mail and Internet Surveys: The Tailored Design Method (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Dimitrova, Vania / McCalla, Gordon I. / Bull, Susan. (eds.). 2007. Open Learner Models: Future Research Directions. International Journal of Artificial Intelligence in Education. 17(3). 2007.

Dirkin, Kathryn H. / Mishra, Punyashloke / Altermatt, Ellen. 2005. “All Or Nothing: Levels of Sociability of a Pedagogical Software Agent and Its Impact on Student Perceptions and Learning.” In: Journal of Educational Multimedia and Hypermedia. 14(2). April 2005. pp. 113–127.

DiSalvo, Carl / Gemperle, Francine. 2003. “From Seduction to Fulfillment: The Use of Anthropomorphic Form in Design.” In: Proceedings of the International Conference on Designing Pleasurable Products and Interfaces (DPPI 2003). Pittsburgh, Pennsylvania, USA. 23–26 June 2003.

Don, Abbe. 1990. “Narrative and the Interface.” In: Laurel. (ed.). 1990a. pp. 383–391.

Don, Abbe / Brennan, Susan / Laurel, Brenda / Shneiderman, Ben. 1992. “Anthropomorphism: From Eliza to Terminator 2.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI ’92). Monterey, California, USA. 3–7 May 1992.

Donath, Judith. 2000. “Being Real: Questions of Tele-Identity.” In: Goldberg. (ed.). 2000. pp. 296–311.

Donath, Judith. 2001. “Mediated Faces.” In: Proceedings of the 4th International Conference on Cognitive Technology: Instruments of Mind (CT 2001). Coventry, United Kingdom. 6–9 August 2001.

Dourish, Paul. 2001, 2004. “Where the Action Is: The Foundations of Embodied Interaction.” 1st Paperback Edition. Cambridge, Massachusetts: The MIT Press.

Dowling, Carolyn. 2000. "Intelligent Pedagogical Agents in Online Learning Environments." In: Proceedings of the International Conference on Educational Uses of Information and Communication Technologies (ICEUT 2000). 16th IFIP Word Computer Congress. Beijing, China. 21–25 August 2000.

Dowling, Carolyn. 2002. "Software Agents and the 'Human Factor' in the Electronic Classroom." In: Proceedings of the Australian Society for Educational Technology Conference (ASET 2002). Melbourne, Australia. 7–10 July 2002.

Downes, Edward J. / McMillan, Sally J. 2000. "Defining Interactivity: A Qualitative Identification of Key Dimensions." In: New Media & Society. 2(2). June 2000. pp. 157–179.

Downes, Stephen. 2005. "E-Learning 2.0." In: eLearn Magazine. 17 October 2005. Available Online. URL: http://elearnmag.org/subpage.cfm?section=articles&article=29-1. Last Access: 6 September 2009.

Doyle, Patrick O. 1999. "When Is a Communicative Agent a Good Idea?." In: Proceedings of the Workshop on Communicative Agents. 3rd International Conference on Autonomous Agents (Agents '99). Seattle, Washington, USA. 1–5 May 1999.

Doyle, Patrick O. 2004. "Annotated Worlds for Animate Characters." Ph.D. Thesis. Department of Computer Science, Stanford University. Stanford, California, USA. March 2004. Available Online. URL: http://www-cs-students.stanford.edu/~pdoyle/thesis-1.pdf. Last Access: 7 September 2009.

Doyle, Patrick O. / Hayes-Roth, Barbara. 1998. "Agents in Annotated Worlds." In: Proceedings of the 2nd International Conference on Autonomous Agents (Agents '98). Minneapolis, Minnesota, USA. 10–13 May 1998.

Draper, Stephen W. 2006. "The Hawthorne, Pygmalion, Placebo and Other Effects of Expectation: Some Notes." Department of Psychology, University of Glasgow. Glasgow, United Kingdom. Available Online. URL: http://www.psy.gla.ac.uk/~steve/hawth.html. Last Access: 7 September 2009.

Drave, William A. 2000. Teaching Online. River Falls, Wisconsin: LERN Books.

Driscoll, Marcy P. 2000. Psychology of Learning for Instruction (2nd ed.). Needham Heights, Maryland: Allyn & Bacon.

Duffy, Brian R. 2003. "Anthropomorphism and the Social Robot." In: Robotics and Autonomous Systems. 42(3–4). March 2003. pp. 177–190.

Duffy, Thomas M. / Cunningham, Donald J. 1996. "Constructivism: Implications for the Design and Delivery of Instruction." In: Jonassen. (ed.). 1996. pp. 170–198.

Duffy, Thomas M. / Jonassen, David H. (eds.). 1992. Constructivism and the Technology of Instruction: A Conversation. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Duit, Reinders. 1995. "The Constructivist View: A Fashionable and Fruitful Paradigm for Science Education Research and Practice." In: Steffe & Gale. (eds.). 1995. pp. 271–285.

Duncan, Starkey, Jr. 1974. "On the Structure of Speaker-Auditor Interaction during Speaking Turns." In: Language in Society. 3(2). October 1974. pp. 161–180.

Durlak, Jerome T. 1987. "A Typology for Interactive Media." In: McLaughlin. (ed.). 1987. pp. 743–757.

Dutoit, Thierry. 1997. An Introduction to Text-to-Speech Synthesis. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dutoit, Thierry / Stylianou, Yannis. 2003. "Text-to-Speech Synthesis." In: Mitkov. (ed.). 2003. pp. 323–338.

Dyer, Michael G. / Cullingford, Richard E. / Alvarado, Sergio J. 1992. “Scripts.” In: Shapiro. (ed.). 1992. Volume 2. pp. 1443–1460.

Dziuban, Charles D. / Hartman, Joel L. / Moskal, Patsy D. 2004. “Blended Learning.” EDUCAUSE Research Bulletin. 2004(7). 30 March 2004. Available Online. URL: http://www.educause.edu/ir/library/pdf/ERB0407.pdf. Last Access: 6 September 2009.

Earnshaw, Rae A. / Guedj, Richard A. / van Dam, Andries / Vince, John A. (eds.). 2001. Frontiers of Human-Centred Computing, Online Communities and Virtual Environments. London: Springer Verlag.

Ebert, Roger. 2004. “Team America: World Police.” Movie Review. In: Chicago Sun-Times. 15 October 2004. Available Online. URL: http://rogerebert.suntimes.com/apps/pbcs.dll/article?AID=/20041014/REVIEWS/40921007. Last Access: 6 September 2009.

Eco, Umberto. 1979. A Theory of Semiotics. Bloomington: Indiana University Press.

Edmondson, R. Shawn. 2005. “Evaluating the Effectiveness of a Telepresence-Enabled Cognitive Apprenticeship Model of Teacher Professional Development.” Ph.D. Thesis. Utah State University. Logan, Utah, USA. 2005. Available Online. URL: http://citeseerx.ksu.edu.sa/viewdoc/summary?doi=10.1.1.101.2404. Last Access: 6 September 2009.

Edmunds, Angela / Morris, Anne. 2000. “The Problem of Information Overload in Business Organizations: A Review of the Literature.” In: International Journal of Information Management. 20(1). February 2000. pp. 17–28.

Egges, Arjan / Kshirsagar, Sumedha / Magnenat-Thalmann, Nadia. 2004. “Generic Personality and Emotion Simulation for Conversational Agents.” In: Computer Animation and Virtual Worlds. 15(1). March 2004. pp. 1–13.

Ehsani, Farzad / Knodt, Eva. 1998. “Speech Technology in Computer-Aided Language Learning: Strengths and Limitations of a New CALL Paradigm.” In: Language Learning & Technology. 2(1). July 1998. pp. 54–73.

Eid, Mohamad / Artail, Hassan / Kayssi, Ayman / Chehab, Ali. 2005. “Trends in Mobile Agent Applications.” In: Journal of Research and Practice in Information Technology. 37(4). November 2005. pp. 323–351.

Eklund, Robert. 2004. “Disfluency in Swedish Human-Human and Human-Machine Travel Booking Dialogues.” Ph.D. Thesis. Linköping Studies in Science and Technology, Department of Computer and Information Science, Linköping University. Linköping, Sweden. Dissertation No. 882. 2004. Available Online. URL: http://www.ep.liu.se/smash/get/diva2:20923/FULLTEXT01. Last Access: 6 September 2009.

Ekman, Paul. 1979. “About Brows: Emotional and Conversational Signals.” In: Von Cranach et al. (eds.). 1979. pp. 169–202.

Ekman, Paul. 1982. Emotions in the Human Face (2nd ed.). Cambridge: Cambridge University Press.

Ekman, Paul. 1984. “Expression and the Nature of Emotion.” In: Scherer & Ekman. (eds.). 1984. pp. 319–344.

Ekman, Paul. 1992a. “An Argument for Basic Emotions.” In: Cognition & Emotion. 6(3–4). May 1992. pp. 169–200.

Ekman, Paul. 1992b. “Are There Basic Emotions?.” In: Psychological Review. 99(3). July 1992. pp. 550–553.

Ekman, Paul. 1999a. “Basic Emotions.” In: Dalgleish & Power. (eds.). 1999. pp. 45–60.

Ekman, Paul. 1999b. “Facial Expressions.” In: Dalgleish & Power. (eds.). 1999. pp. 301–320.

Ekman, Paul / Friesen, Wallace V. 1969. “The Repertoire of Nonverbal Behaviors: Categories, Origins, Usage, and Coding.” In: Semiotica. 1. 1969. pp. 49–98.

Ekman, Paul / Friesen, Wallace V. 1975. Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues. Englewood Cliffs, New Jersey: Prentice Hall.

Ekman, Paul / Friesen, Wallace V. 1978. Facial Action Coding System. Palo Alto, California: Consulting Psychologists Press.

Ekman, Paul / Friesen, Wallace V. / Hager, Joseph C. 2002. Facial Action Coding System (2nd ed.). London: Weidenfeld & Nicholson.

Ekman, Paul / Rosenberg, Erika L. (eds.). 2005. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (2nd ed.). New York: Oxford University Press.

Elliott, Clark. 2003. “The Role of Elegance in Emotion and Personality: Reasoning for Believable Agents.” In: Trappl et al. (eds.). 2003. pp. 237–249.

Elliott, Clark / Brzezinski, Jacek. 1998. “Autonomous Agents as Synthetic Characters.” In: AI Magazine. 19(2). Summer 1998. pp. 13–30.

Elliott, Clark / Rickel, Jeff W. / Lester, James C. 1999. “Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis.” In: Wooldridge & Veloso. (eds.). 1999. pp. 195–212.

Ellsworth, Phoebe C. / Scherer, Klaus R. 2003. “Appraisal Processes in Emotion.” In: Davidson et al. (eds.). 2003. pp. 572–595.

Emmerling, Robert J. / Goleman, Daniel. 2003. “Emotional Intelligence: Issues and Common Misunderstandings.” In: Issues and Recent Developments in Emotional Intelligence. 1(1). October 2003. Available Online. URL: http://www.eiconsortium.org/pdf/EI_Issues_And_Common_Misunderstandings.pdf. Last Access: 6 September 2009.

Engelmore, Robert S. / Morgan, Anthony. (eds.). 1988. Blackboard Systems. London: Addison-Wesley.

Enos, Frank. 2003. “A Survey of Models of Emotion Useful in the Context of Computation: Annotated Bibliography.” Columbia University. New York City, New York, USA. 1 September 2003. Available Online. URL: http://www.cs.columbia.edu/~frank/papers/emotion_annotated.pdf. Last Access: 6 September 2009.

Eppler, Martin J. / Mengis, Jeanne. 2004. “The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines.” In: The Information Society. 20(5). November–December 2004. pp. 325–344.

Epstein, Robert / Roberts, Gary / Beber, Grace. (eds.). 2008. Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. New York: Springer Verlag.

Ercoli, Paolo / Lewis, Robert. (eds.). 1988. Artificial Intelligence Tools in Education: Proceedings of the IFIP TC 3 Working Conference on Artificial Intelligence Tools in Education. Frascali, Italy. 26–28 May 1987. New York: Elsevier Science.

Erickson, Thomas. 1997. “Designing Agents as if People Mattered.” In: Bradshaw. (ed.). pp. 79–96.

Ericsson, K. Anders / Simon, Herbert A. 1984. Protocol Analysis: Verbal Report as Data. Cambridge, Massachusetts: The MIT Press.

Ertmer, Peggy A. / Newby, Timothy J. 1993. “Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective.” In: Performance Improvement Quarterly. 6(4). December 1993. pp. 50–72.

Eskenazi, Maxine. 1999. “Using Automatic Speech Recognition for Foreign Language Pronunciation Tutoring: Some Issues and a Prototype.” In: Language Learning & Technology. 2(2). January 1999. pp. 62–76.

Euler, Dieter. 1994. “(Multi)Mediales Lernen: Theoretische Fundierung und Forschungsstand.” In: Unterrichtswissenschaft. 22(4). 1994. pp. 291–311.

Evans, Dylan. 2004. “Can Robots Have Emotions?.” In: Psychology Review. 11(1). September 2004. pp. 2–5.

Faber, Wolfgang. 1993. “Hypermediale Lernsysteme.” In: Proceedings of the Conference “Forschung für die Wirtschaft: Im Mittelpunkt: der Mensch” (Annual Conference of the Vienna University of Economics and Business). Vienna, Austria. 11–12 March 1993. Available Online. URL: http://wwwai.wu-wien.ac.at/Publikationen/Faber/WU-JT.html. Last Access: 6 September 2009.

Fayyad, Usama / Piatetsky-Shapiro, Gregory / Smyth, Padhraic. 1996. “From Data Mining to Knowledge Discovery in Databases.” In: AI Magazine. 17(3). Fall 1996. pp. 37–54.

Feldman, Robert S. / Rimé, Bernard. (eds.). 1991. Fundamentals of Nonverbal Behavior. Cambridge: Cambridge University Press.

Fellbaum, Klaus. 1997. “Sprach- und Sprechererkennung: Prinzipien, Anwendungen und Stand der Technik.” In: Spektrum der Wissenschaft, Dossier: “Kopf oder Computer.” Dossier 1997(4). October 1997. pp. 42–47.

Fellous, Jean-Marc / Arbib, Michael A. (eds.). 2005. Who Needs Emotions?: The Brain Meets the Robot. New York: Oxford University Press.

Fensel, Dieter / Hendler, James A. / Lieberman, Henry / Wahlster, Wolfgang. (eds.). 2003. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. Cambridge, Massachusetts: The MIT Press.

Fensel, Dieter / Musen, Mark A. 2001. “The Semantic Web: A Brain for Humankind.” In: IEEE Intelligent Systems. 16(2). March–April 2001. pp. 24–25.

Ferdig, Richard E. / Mishra, Punyashloke. 2004. “Emotional Responses to Computers: Experiences in Unfairness, Anger, and Spite.” In: Journal of Educational Multimedia and Hypermedia. 13(2). April 2004. pp. 143–161.

Ferguson, George / Allen, James F. 1998. “TRIPS: An Integrated Intelligent Problem-Solving Assistant.” In: Proceedings of the 15th National Conference on Artificial Intelligence (AAAI ’98). Madison, Wisconsin, USA. 26–30 July 1998.

Ferguson, George / Allen, James F. 2007. “Mixed-Initiative Dialogue Systems for Collaborative Problem-Solving.” In: AI Magazine. 28(2). Summer 2007. pp. 23–32.

Fern, Alan. 2007. “Speedup Learning.” Department of Computer Science, Oregon State University. Corvallis, Oregon, USA. 28 June 2007. Available Online. URL: http://web.engr.oregonstate.edu/~afern/papers/speedup.pdf. Last Access: 6 September 2009.

Fernández-Dols, José-Miguel / Carroll, James M. 1997. “Is the Meaning Perceived in Facial Expression Independent of Its Context?.” In: Russell & Fernández-Dols. (eds.). 1997. pp. 275–294.

Finin, Timothy W. / Fritzson, Richard / McKay, Don / McEntire, Robin. 1994. “KQML as an Agent Communication Language.” In: Proceedings of the 3rd International Conference on Information and

Knowledge Management (CIKM ’94). Gaithersburg, Maryland, USA. 29 November–02 December 1994.

Finin, Timothy W. / Labrou, Yannis / Mayfield, James. 1997. “KQML as an Agent Communication Language.” In: Bradshaw. (ed.). 1997. pp. 291–316.

Fletcher, J. Dexter / Tobias, Sigmund. 2005. “The Multimedia Principle.” In: Mayer. (ed.). 2005a. pp. 117–133.

Flottemesch, Kim. 2000. “Building Effective Interaction in Distance Education: A Review of the Literature.” In: Educational Technology. 40(3). May–June 2000. pp. 46–51.

Fogg, Brian J. 1997. “Charismatic Computers: Creating More Likable and Persuasive Interactive Technologies by Leveraging Principles from Social Psychology.” Ph.D. Thesis. Department of Communication, Stanford University. Stanford, California, USA. June 1997.

Fogg, Brian J. 2003. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco: Morgan Kaufmann Publishers.

Fogg, Brian J. / Nass, Clifford I. 1997. “Silicon Sycophants: The Effects of Computers that Flatter.” In: International Journal of Human-Computer Studies. 46(5). May 1997. pp. 551–561.

Foner, Leonard N. 1993. “What’s an Agent Anyway?: A Sociological Case Study.” Agents Memo 93-01. Software Agents Group, Media Laboratory, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. May 1993. Available Online. URL: http://pubs.media.mit.edu/pubs/papers/Agents--Julia.pdf. Last Access: 7 September 2009.

Fong, Terrence / Nourbakhsh, Illah / Dautenhahn, Kerstin. 2003. “A Survey of Socially Interactive Robots.” In: Robotics and Autonomous Systems. 42(3–4). 31 March 2003. pp. 143–166.

Forbus, Kenneth D. / Feltovich, Paul. J. (eds.). 2001. Smart Machines in Education. Cambridge, Massachusetts: The MIT Press.

Forsyth, David A. / Ponce, Jean. 2003. Computer Vision: A Modern Approach. Upper Saddle River, New Jersey: Prentice Hall.

Forsythe, Chris / Bernard, Michael L. / Goldsmith, Timothy E. (eds.). 2005. Cognitive Systems: Human Cognitive Models in Systems Design. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Franke, Peter. 2003. “Prinzipien des E-Learning im VLC: Blended und Online.” In: Proceedings of the 7th Workshop on Multimedia for Education and Economy (“Multimedia für Bildung und Wirtschaft”). Ilmenau University of Technology. Ilmenau, Germany. 25–26 September 2003.

Franke, Peter. 2006. “Future Developments: Designing Embodied Pedagogical Agents for the VLC.” In: Handke & Franke. (eds.). 2006. pp. 225–311.

Franke, Peter / Handke, Jürgen / Hente, Astrid / Sperl, Alexander / Unger, Tobias. 2006. “The VLC E-Learning Concept.” The Virtual Linguistics Campus (VLC). Philipps University. Marburg, Germany. Available Online. URL: http://linguistics.online.uni-marburg.de/free/information/portal/elearning_concept.pdf. Last Access: 7 September 2009.

Frankle, Kathleen M. 2005. “Blended Learning: The Key to Successful Web-Based Training and Education.” In: Proceedings of the 12th World Congress on Intelligent Transport Systems (ITS 2005). San Francisco, California, USA. 6–10 November 2005.

Franklin, James. 2003. “The Representation of Context: Ideas from Artificial Intelligence.” In: Law, Probability and Risk. 2(3). September 2003. pp. 191–199.

Franklin, Stan / Graesser, Arthur C. 1996. “Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents.” In: Proceedings of the 3rd International Workshop on Agent Theories, Architectures, and Languages (ATAL ’96). Budapest, Hungary. 12–13 August 1996.

Frasson, Claude / Mengelle, Thierry / Aïmeur, Esma. 1997. “Using Pedagogical Agents in a Multi-Strategic Intelligent Tutoring System.” In: Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED ’97). Kobe, Japan. 18–22 August 1997.

Frasson, Claude / Gauthier, Gilles. (eds.). 1990. Intelligent Tutoring Systems: At the Crossroads of Artificial Intelligence and Education. Norwood, New Jersey: Ablex Publishing.

Frawley, William J. / Piatetsky-Shapiro, Gregory / Matheus, Christopher J. 1992. “Knowledge Discovery in Databases: An Overview.” In: AI Magazine. 13(3). Fall 1992. pp. 57–70.

Freedman, Reva. 2000. “What Is an Intelligent Tutoring System?.” In: Intelligence. 11(3). September 2000. pp. 15–16.

Friedland, Carsten. 2001. “The Application of Information and Communication Technologies in Learning and Training in Developing Countries.” M.Sc. Thesis. Karlsruhe University of Applied Sciences. Karlsruhe, Germany. 26 September 2001. Available Online. URL: http://hagar.up.ac.za/catts/learner/friedland/Thesis_Friedland_26.09.2001.pdf. Last Access: 7 September 2009.

Friedman, Bayta. (ed.). 1997. Human Values and the Design of Computer Technology. Stanford, California: CSLI Publications.

Frijda, Nico H. 1986. The Emotions. Cambridge: Cambridge University Press.

Frijda, Nico H. 1993. “The Place of Appraisal in Emotion.” In: Cognition & Emotion. 7(3–4). May 1993. pp. 357–387.

Fulk, Janet / Steinfeld, Charles W. (eds.). 1990. Organization and Communication Technology. Newbury Park, California: SAGE Publications.

Gagné, Robert M. / Wager, Walter W. / Golas, Katherine C. / Keller, John M. 2004. Principles of Instructional Design (5th ed.). Belmont, California: Wadsworth Publishing.

Garau, Maia / Slater, Mel / Vinayagamoorthy, Vinoba / Brogni, Andrea / Steed, Anthony / Sasse, M. Angela. 2003. “The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2003). Fort Lauderdale, Florida, USA. 5–10 April 2003.

Gard, Toby. 2000. “Building Character.” Gama Network, Gamasutra.com. 20 July 2000. No Longer Available Online. URL: http://www.gamasutra.com/features/20000720/gard_pfv.htm. Last Access: 22 November 2008.

Garrison, D. Randy. 1993. “Quality and Access in Distance Education: Theoretical Considerations.” In: Keegan. (ed.). 1993. pp. 9–21.

Gasser, Leslie G. 1989. “Distributed Artificial Intelligence.” In: AI Expert. 4(7). July 1989. pp. 26–33.

Gavrila, Dariu M. 1999. “The Visual Analysis of Human Movement: A Survey.” In: Computer Vision and Image Understanding. 73(1). January 1999. pp. 82–98.

Gazdar, Gerald. 1993. “The Handling of Natural Language.” In: Broadbent. (ed.). 1993. pp. 151–177.

Gazdar, Gerald / Mellish, Christopher S. 1989. Natural Language Processing in PROLOG: An Introduction to Computational Linguistics. New York: Addison-Wesley.

Gebhard, Patrick / Kipp, Michael / Klesen, Martin / Rist, Thomas. 2003. "What Are They Going to Talk About?: Towards Life-Like Characters that Reflect on Interactions with Users." In: Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2003). Darmstadt, Germany. 24–26 March 2003.

Gee, James P. 2003. What Video Games Have to Teach Us About Learning and Literacy. New York: Palgrave Macmillan.

Genesereth, Michael R. / Fikes, Richard E. 1992. "Knowledge Interchange Format Version 3.0 Reference Manual." Technical Report Logic-92-1. Logic Group, Computer Science Department, Stanford University. Stanford, California, USA. June 1992. Available Online. URL: http://www-ksl.stanford.edu/knowledge-sharing/papers/kif.ps. Last Access: 7 September 2009.

Genesereth, Michael R. / Ketchpel, Steven P. 1994. "Software Agents." In: Communications of the ACM. 37(7). July 1994. pp. 48–53.

Gentner, Dedre / Stevens, Albert L. (eds.). 1983. Mental Models. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Georgeff, Michael / Pell, Barney / Pollack, Martha / Tambe, Milind / Wooldridge, Michael J. 1998. "The Belief-Desire-Intention Model of Agency." In: Proceedings of the 5th International Workshop on Agent Theories, Architectures, and Languages (ATAL '98). Paris, France. 4–7 July 1998.

Gertner, Abigail S. 1999. "Some Challenges for the Use of Probabilities in a Student Modeling Standard." In: Proceedings of the 7th International Conference on User Modeling (UM '99). Banff, Alberta, Canada. 20–24 June 1999.

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?." In: Analysis. 23(6). June 1963. pp. 121–123. Available Online. URL: http://www.ditext.com/gettier/gettier.html. Last Access: 7 September 2009.

Ghezzi, Carlo / Jazayeri, Mehdi / Mandrioli, Dino. 2002. Fundamentals of Software Engineering (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Giachin, Egidio. 1997. "Spoken Language Dialogue." In: Cole et al. (eds.). 1997. pp. 210–214.

Gibbs, Anita. 1997. "Focus Groups." In: Social Research Update. Issue 19. Winter 1997. Available Online. URL: http://sru.soc.surrey.ac.uk/SRU19.html. Last Access: 7 September 2009.

Gibson, Chère C. (ed.). 1998a. Distance Learners in Higher Education: Institutional Responses for Quality Outcomes. Madison, Wisconsin: Atwood Publishing.

Gibson, Chère C. 1998b. "The Distance Learner in Context." In: Gibson. (ed.). 1998a. pp. 113–125.

Gilbert, Larry / Moore, David R. 1998. "Building Interactivity into Web Courses: Tools for Social and Instructional Interaction." In: Educational Technology. 38(3). May–June 1998. pp. 29–35.

Glaser, Barney G. 1992. Basics of Grounded Theory Analysis. Mill Valley, California: Sociology Press.

Glasow, Priscilla A. 2005. "Fundamentals of Survey Research Methodology." Technical Paper 05-0638. The MITRE Corporation. Mclean, Virginia, USA. April 2005. Available Online. URL: http://www.mitre.org/work/tech_papers/tech_papers_05/05_0638/05_0638.pdf. Last Access: 7 September 2009.

Godwin, Steve. 2005. "A Preliminary Evaluation of the Concepts of Interaction and Interactivity in Computer Mediated Learning Environments." Research Project Report KN6083. The Open University. Milton Keynes, United Kingdom. 31 March 2005. Available Online. URL: http://kn.open.ac.uk/public/getfile.cfm?documentfileid=6093. Last Access: 7 September 2009.

Goffman, Erving. 1955. "On Face-Work: An Analysis of Ritual Elements in Social Interaction." In: Psychiatry: Journal of the Study of Interpersonal Processes. 18(3). August 1955. pp. 213–231.

Goffman, Erving. 1967. Interaction Ritual: Essays on Face-to-Face Behavior. New York: Pantheon Books.

Goh, Ong Sing / Fung, Chun Che / Wong, Kok Wai / Depickere, Arnold. 2006. "An Embodied Conversational Agent for Intelligent Web Interaction on Pandemic Crisis Communication." In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006). Hong Kong, China. 18–22 December 2006.

Golbeck, Jennifer / Hendler, James. 2004. "Inferring Reputation on the Semantic Web." In: Proceedings of the 13th International World Wide Web Conference (WWW 2004). New York City, New York, USA. 17–22 May 2004.

Golbeck, Jennifer / Parsia, Bijan / Hendler, James. 2003. "Trust Networks on the Semantic Web." In: Proceedings of the 7th International Workshop on Cooperative Information Agents (CIA 2003). Helsinki, Finland. 27–29 August 2003.

Golbeck, Jennifer / Alford, Amy / Alford, Ron / Hendler, James. 2005. "Organization and Structure of Information Using Semantic Web Technologies." In: Proctor & Vu. (eds.). 2005. pp. 176–189.

Goldberg, Ken. (ed.). 2000. The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet. Cambridge, Massachusetts: The MIT Press.

Goldstein, Ira P. 1982. "The Genetic Graph: A Representation for the Evolution of Procedural Knowledge." In: Sleeman & Brown. (eds.). 1982. pp. 51–77.

Goldberg, Lewis R. 1990. "An Alternative 'Description of Personality:' The Big-Five Factor Structure." In: Journal of Personality and Social Psychology. 59(6). December 1990. pp. 1216–1229.

Goldberg, Lewis R. / Saucier, Gerard 1998. "What Is Beyond the Big Five?." In: Journal of Personality. 66(4). August 1998. pp. 495–524.

Gombrich, Ernst H. 1972. Art and Illusion: A Study in the Psychology of Pictorial Representation (4th ed.). London: Phaidon Press.

Gong, Li. 2000. "The Psychology of Consistency in Human-Computer Interaction." Ph.D. Thesis. Department of Communication, Stanford University. Stanford, California, USA. 2000.

Gong, Li. 2007. "Is Happy Better Than Sad Even If They Are Both Non-Adaptive?: Effects of Emotional Expressions of Talking-Head Interface Agents." In: International Journal of Human-Computer Studies. 65(3). March 2007. pp. 183–191.

Gong, Li / Lai, Jennifer. 2001. "Shall We Mix Synthetic Speech and Human Speech?: The Impact on Users' Task Performance and Attitude." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2001). Seattle, Washington, USA. 31 March–5 April 2001.

Gong, Li / Nass, Clifford I. 2000. "Does Adding a Synthetic Face Always Enhance Speech Interfaces?." Department of Communication, Stanford University. Stanford, California, USA. Available Online. URL: http://www.stanford.edu/~nass/comm369/pdf/SyntheticFaceandTypeOfVoice.pdf. Last Access: 7 September 2009.

Gong, Li / Nass, Clifford I. 2007. "When a Talking-Face Computer Agent Is Half-Human and Half-Humanoid: Human Identity and Consistency Preference." In: Human Communication Research. 33(2). April 2007. pp. 163–193.

Gong, Li / Nass, Clifford I. / Takhteyev, Yuri V. 2001. "When Non-Human Is Better than Semi-Human: Consistency in Speech Interfaces." In: Proceedings of the 9th International Conference on Human-Computer Interaction (HCI 2001). New Orleans, Louisiana, USA. 5–10 August 2001.

González, Carolina / Burguillo, Juan C. / Llamas, Martin. 2006. "A Qualitative Comparison of Techniques for Student Modeling in Intelligent Tutoring Systems." In: Proceedings of the 36th Annual Frontiers in Education Conference (FIE 2006). San Diego, California, USA. 28–31 October 2006.

Good, Ron / Berger, Carl. 1998. "The Computer as a Powerful Tool for Understanding Science." In: Mintzes et al. (eds.). 1998. pp. 213–227.

Goodwin, Charles. 1981. Conversational Organization: Interaction Between Speakers and Hearers. New York: Academic Press.

Goodyear, Peter. (ed.). 1991. Teaching Knowledge and Intelligent Tutoring. Norwood, New Jersey: Ablex Publishing.

Gorham, Joan. 1988. "The Relationship Between Verbal Teacher Immediacy Behaviors and Student Learning." In: Communication Education. 37(1). January 1988. pp. 40–53.

Graesser, Arthur C. / D'Mello, Sidney K. / Craig, Scotty D. / Witherspoon, Amy / Sullins, Jeremiah / McDaniel, Bethany / Gholson, Barry. 2008. "The Relationship Between Affective States and Dialog Patterns during Interactions with AutoTutor." In: Journal of Interactive Learning Research. 19(2). April 2008. pp. 293–312.

Graesser, Arthur C. / Jackson, G. Tanner / McDaniel, Bethany. 2007. "AutoTutor Holds Conversations with Learners that Are Responsive to Their Cognitive and Emotional States." In: Educational Technology. 47(1). January–February 2007. pp. 19–22.

Graesser, Arthur C. / McDaniel, Bethany / Chipman, Patrick / Witherspoon, Amy / D'Mello, Sidney K. / Gholson, Barry. 2006. "Detection of Emotions during Learning with AutoTutor." In: Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006). Vancouver, British Columbia, Canada. 26–29 July 2006.

Graesser, Arthur C. / Olney, Andrew / Haynes, Brian C. / Chipman, Patrick. 2005. "AutoTutor: A Cognitive System that Simulates a Tutor that Facilitates Learning Through Mixed-Initiative Dialogue." In: Forsythe et al. (eds.). 2005. pp. 177–212.

Graesser, Arthur C. / Person, Natalie K. / Magliano, Joseph P. 1995. "Collaborative Dialogue Patterns in Naturalistic One-to-One Tutoring Sessions." In: Applied Cognitive Psychology. 9(6). December 1995. pp. 495–522.

Graesser, Arthur C. / VanLehn, Kurt / Rose, Carolyn / Jordan, Pamela / Harter, Derek. 2001. "Intelligent Tutoring Systems with Conversational Dialogue." In: AI Magazine. 22(4). Winter 2001. pp. 39–51.

Grand, Stephen / Cliff, Dave. 1998. "Creatures: Entertainment Software Agents with Artificial Life." In: Autonomous Agents and Multi-Agent Systems. 1(1). March 1998. pp. 39–57.

Grand, Stephen / Cliff, Dave / Malhotra, Anil. 1997. "Creatures: Artificial Life Autonomous Software Agents for Home Entertainment." In: Proceedings of the 1st International Conference on Autonomous Agents (Agents '97). Marina del Rey, California, USA. 5–8 February 1997.

Grandison, Tyrone / Sloman, Morris. 2000. "A Survey of Trust in Internet Applications." In: IEEE Communications Surveys. 3(4). September 2000. pp. 2–16.

Gratch, Jonathan / Marsella, Stacy. 2003. "Fight the Way You Train: The Role and Limits of Emotions in Training for Combat." In: The Brown Journal of World Affairs. 10(1). Summer–Fall 2003. pp. 63–76.

Gratch, Jonathan / Marsella, Stacy. 2004. “A Domain-Independent Framework for Modeling Emotion.” In: Journal of Cognitive Systems Research. 5(4). December 2004. pp. 269–306.

Gratch, Jonathan / Marsella, Stacy. 2005a. “Lessons from Emotion Psychology for the Design of Lifelike Characters.” In: Applied Artificial Intelligence. 19(3–4). March 2005. pp. 215–233.

Gratch, Jonathan / Marsella, Stacy. 2005b. “Evaluating a Computational Model of Emotion.” In: Autonomous Agents and Multi-Agent Systems. 11(1). July 2005. pp. 23–43.

Gratch, Jonathan / Rickel, Jeff W. / André, Elisabeth / Badler, Norman I. / Cassell, Justine / Petajan, Eric. 2002. “Creating Interactive Virtual Humans: Some Assembly Required.” In: IEEE Intelligent Systems. 17(4). July–August 2002. pp. 54–63.

Gray, Chris H. / Figueroa-Sarriera, Heidi J. / Mentor, Steven. (eds.). 1995. The Cyborg Handbook. New York: Routledge.

Green, Robert D. / MacDorman, Karl F. / Ho, Chin-Chang / Vasudevan, Sandosh. 2008. “Sensitivity to the Proportions of Faces that Vary in Human-Likeness.” In: Computers in Human Behavior. 24(5). September 2008. pp. 2456–2474.

Greenfield, Patricia M. 1984. “A Theory of the Teacher in the Learning Activities of Everyday Life.” In: Rogoff & Lave. (eds.). 1984. pp. 117–138.

Greer, Jim E. / McCalla, Gordon I. (eds.). 1994. Student Modelling: The Key to Individualized Knowledge-Based Instruction. Berlin: Springer Verlag.

Gregory, Richard L. (ed.). 2004. The Oxford Companion to the Mind (2nd ed.). Oxford: Oxford University Press.

Greve, Steven H. 1990. “Proactive Modelling and the Development of Coaching Strategies in Geometry.” In: Proceedings of the 3rd International Conference on Computer-Assisted Learning (ICCAL ’90). Hagen, Germany. 11–13 June 1990.

Grice, Herbert P. 1975. “Logic and Conversation.” In: Cole & Morgan. (eds.). 1975. pp. 41–58.

Grishman, Ralph. 2003. “Information Extraction.” In: Mitkov. (ed.). 2003. pp. 545–559.

Grob, Heinz Lothar / Seufert, Sabine. 1996. “Erwartungen der Praxis an CAL – Ergebnisse einer Expertenbefragung.” Research Report. Department of Business Informatics, University of Münster. Münster, Germany. July 1996. Available Online. URL: http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-1591/CALCAT06.pdf. Last Access: 7 September 2009.

Grossman, David A. / Frieder, Ophir. 2004. Information Retrieval: Algorithms and Heuristics (2nd ed.). Dordrecht, The Netherlands: Springer Verlag.

Grosz, Barbara J. 1997. “Discourse and Dialogue: Overview.” In: Cole et al. (eds.). 1997. pp. 199–201.

Grosz, Barbara J. / Sidner, Candace L. 1986. “Attention, Intention, and the Structure of Discourse.” In: Computational Linguistics. 12(3). July–September 1986. pp. 175–204.

Gruber, Thomas R. 1993. “A Translation Approach to Portable Ontology Specifications.” In: Knowledge Acquisition. 5(2). June 1993. pp. 199–220.

Gudykunst, William B. / Kim, Young Yun. 1992. Communicating with Strangers: An Approach to Intercultural Communication (2nd ed.). New York: McGraw-Hill.

Guedj, Richard A. / ten Hagen, Paul J. W. / Hopgood, Frank R. A. / Tucker, Hugh A. / Duce, David A. (eds.). 1980. Methodology of Interaction. Proceedings of the IFIP Workshop. Seillac, France. May 1979. Amsterdam: North-Holland Publishing Company.

Gugerty, Leo. 1997. "Non-Diagnostic Intelligent Tutoring Systems: Teaching Without Student Models." In: Instructional Science. 25(6). November 1997. pp. 409–432.

Gulz, Agneta. 2004. "Benefits of Virtual Characters in Computer Based Learning Environments: Claims and Evidence." In: International Journal of Artificial Intelligence in Education. 14(3–4). 2004. pp. 313–334.

Gulz, Agneta / Haake, Magnus. 2006. "Design of Animated Pedagogical Agents: A Look at Their Look." In: International Journal of Human-Computer Studies. 64(4). April 2006. pp. 322–339.

Guri-Rosenblit, Sarah. 2005. "Eight Paradoxes in the Implementation Process of E-Learning in Higher Education." In: Higher Education Policy. 18(1). March 2005. pp. 5–29.

Guthrie, James W. (ed.). 2003. The Encyclopedia of Education (2nd ed.). New York: MacMillan Publishers.

Guy, Marieke / Tonkin, Emma. 2006. "Folksonomies: Tidying up Tags?." In: D-Lib Magazine. 12(1). January 2006. Available Online. URL: http://www.dlib.org/dlib/january06/guy/01guy.html. Last Access: 7 September 2009.

Ha, Louisa / James, E. Lincoln. 1998. "Interactivity Reexamined: A Baseline Analysis of Early Business Web Sites." In: Journal of Broadcasting & Electronic Media. 42(4). Fall 1998. pp. 457–474.

Haake, Magnus / Gulz, Agneta. 2007. "Virtual Pedagogical Agents: Stylisation for Engagement." In: Interfaces. 70. Spring 2007. pp. 12–13. Available Online. URL: http://www.bcs-hci.org.uk/files/usermedia/file/interfaces_69on/interfaces70.pdf. Last Access: 7 September 2009.

Haake, Magnus / Gulz, Agneta. 2008. "Visual Stereotypes and Virtual Pedagogical Agents." In: Educational Technology & Society. 11(4). October 2008. pp. 1–15.

Haddad, Hanadi / Klobas, Jane. 2002. "The Relationship Between Visual Abstraction and the Effectiveness of a Pedagogical Character Agent." In: Proceedings of the Workshop "Embodied Conversational Agents – Let's Specify and Evaluate Them!." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Haddawy, Peter. 1999. "An Overview of Some Recent Developments in Bayesian Problem Solving Techniques." In: AI Magazine. 20(2). Summer 1999. pp. 11–19.

Haeckel, Stephan H. 1998. "About the Nature and Future of Interactive Marketing." In: Journal of Interactive Marketing. 12 (1). Winter 1998. pp. 63–71.

Hall, Edward T. 1976. Beyond Culture. New York: Doubleday.

Hand, David J. / Mannila, Heikki / Smyth, Padhraic. 2001. Principles of Data Mining. Cambridge, Massachusetts: The MIT Press.

Handke, Jürgen. 1995. The Structure of the Lexicon: Human Versus Machine. Berlin: Mouton de Gruyter.

Handke, Jürgen. 2006a. "The Virtual Linguistics Campus – An Overview." In: Handke & Franke. (eds.). 2006. pp. 14–35.

Handke, Jürgen. 2006b. "VLC E-Bologna: Supporting the Bologna Process." In: Handke & Franke. (eds.). 2006. pp. 136–148.

Handke, Jürgen / Franke, Peter. (eds.). 2006. The Virtual Linguistics Campus: Strategies and Concepts for Successful E-Learning. Münster: Waxmann Verlag.

Hansson, Sven O. 1994, 2005. "Decision Theory: A Brief Introduction." Department of Philosophy and the History of Technology, Royal Institute of Technology (KTH). Stockholm, Sweden. 19 August 1994. Minor Revisions 23 August 2005. Available Online. URL: http://www.infra.kth.se/~soh/decisiontheory.pdf. Last Access: 8 September 2009.

Harabagiu, Sanda / Moldovan, Dan. 2003. "Question Answering." In: Mitkov. (ed.). 2003. pp. 560–582.

Hardy, Melissa / Bryman, Alan. (eds.). 2004. Handbook of Data Analysis. London: SAGE Publications.

Harkins, Stephen G. / Petty, Richard E. 1981. "The Effects of Source Magnification on Cognitive Efforts and Attitudes: An Information Processing View." In: Journal of Personality and Social Psychology. 40(3). March 1981. pp. 401–413.

Harkins, Stephen G. / Petty, Richard E. 1983. "Social Context Effects in Persuasion: The Effects of Multiple Sources and Multiple Targets." In: Paulus. (ed.). 1983. pp. 149–175.

Harkins, Stephen G. / Petty, Richard E. 1987. "Information Utility and the Multiple Source Effect in Persuasion." In: Journal of Personality and Social Psychology. 52(2). February 1987. pp. 260–268.

Harnish, Robert M. 2002. Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science. Malden, Massachusetts: Blackwell Publishing.

Hartley, Danita / Mitrovic, Antonija. 2002. "Supporting Learning by Opening the Student Model." In: Proceedings of the 6th International Conference on Intelligent Tutoring Systems (ITS 2002). Biarritz, France and San Sebastian, Spain. 2–7 June 2002.

Hawkins, Robert P. / Wiemann, John M. / Pingree, Suzanne. (eds.). 1988. Advancing Communication Science: Merging Mass and Interpersonal Processes. Newbury Park, California: SAGE Publications.

Hayes-Roth, Barbara. 2004. "What Makes Characters Seem Life-Like?." In: Prendinger & Ishizuka. (eds.). 2004a. pp. 447–462.

Hayes-Roth, Barbara / Doyle, Patrick O. 1998. "Animate Characters." In: Autonomous Agents and Multi-Agent Systems. 1(2). October 1998. pp. 195–230.

Hayes-Roth, Barbara / van Gent, Robert / Huber, Daniel. 1997. "Acting in Character." In: Trappl & Petta. (eds.). 1997. pp. 92–112.

Hayman, Sarah. 2007. "Folksonomies and Tagging: New Developments in Social Bookmarking." In: Proceedings of the Ark Group Conference "Developing and Improving Classification Schemes." Sydney, Australia. 27–29 June 2007.

He, Minghua / Jennings, Nicholas R. / Leung, Ho-Fung. 2003. "On Agent-Mediated Electronic Commerce." In: IEEE Transactions on Knowledge and Data Engineering. 15(4). July–August 2003. pp. 985–1003.

He, Minghua / Leung, Ho-Fung. 2002. "Agents in E-Commerce: State of the Art." In: Knowledge and Information Systems. 4(3). July 2002. pp. 257–282.

Hearst, Marti A. 2003. "Text Data Mining." In: Mitkov. (ed.). 2003. pp. 616–628.

Heckerman, David. 1995, 1996. "A Tutorial on Learning with Bayesian Networks." Technical Report MSR-TR-95-06. Microsoft Research. Redmond, Washington, USA. March 1995. Revised November 1996. Available Online. URL: ftp://ftp.research.microsoft.com/pub/tr/tr-95-06.pdf. Last Access: 8 September 2009.

Heckman, Carey E. / Wobbrock, Jacob O. 2000. "Put Your Best Face Forward: Anthropomorphic Agents, E-Commerce Consumers, and the Law." In: Proceedings of the 4th International Conference on Autonomous Agents (Agents 2000). Barcelona, Spain. 3–7 June 2000.

Heeter, Carrie. 1989. "Implications of Interactivity for Communication Research." In: Salvaggio & Bryant. (eds.). 1989. pp. 217–235.

Heeter, Carrie. 1992. "Being There: The Subjective Experience of Presence." In: Presence: Teleoperators and Virtual Environments. 1(2). Spring 1992. pp. 262–271.

Heeter, Carrie. 2000. "Interactivity in the Context of Designed Experiences." In: Journal of Interactive Advertising. 1(1). Fall 2000. pp. 75–89. Available Online. URL: http://jiad.org/article2. Last Access: 8 September 2009.

Heffner, Christopher L. 2004a. "Experimental Design." In: AllPsych Online: The Virtual Psychology Classroom. Research Methods. Chapter 5. Available Online. URL: http://allpsych.com/researchmethods/experimentaldesign.html. Last Access: 8 September 2009.

Heffner, Christopher L. 2004b. "Inferential Statistics." In: AllPsych Online: The Virtual Psychology Classroom. Research Methods. Chapter 9. Available Online. URL: http://allpsych.com/researchmethods/inferentialstatistics.html. Last Access: 8 September 2009.

Heider, Fritz / Simmel, Marianne. 1944. "An Experimental Study of Apparent Behavior." In: The American Journal of Psychology. 57(2). April 1944. pp. 243–259.

Helander, Martin G. / Landauer, Thomas K. / Pabhu, Prasad V. (eds.). 1997. Handbook of Human-Computer Interaction (2nd ed.). Amsterdam: Elsevier Science.

Helms, James A. / Arthur, James D. / Hix, Deborah / Hartson, H. Rex. 2006. "A Field Study of the Wheel: A Usability Engineering Process Model." In: The Journal of Systems and Software. 79(6). June 2006. pp. 841–858.

Hendler, James. 2001. "Agents and the Semantic Web." In: IEEE Intelligent Systems. 16(2). March–April 2001. pp. 30–37.

Hendler, James. 2008. "Web 3.0: Chicken Farms on the Semantic Web." In: Computer. 41(1). January 2008. pp. 106–108.

Hendler, James / Berners-Lee, Tim / Miller, Eric. 2002. "Integrating Applications on the Semantic Web." In: Journal of the Institute of Electrical Engineers of Japan. 122(10). October 2002. pp. 676–680.

Hente, Astrid. 2006. "Teaching on the VLC." In: Handke & Franke. (eds.). 2006. pp. 67–84.

Hess, Wolfgang. 1996. "Neuere Entwicklungen in der Sprachsynthese." Verbmobil Report No. 174. University of Bonn. Bonn, Germany. September 1996. Available Online. URL: http://www.ims.uni-stuttgart.de/projekte/verbmobil/vm-reports/report-174-96.ps.tar.gz. Last Access: 8 September 2009.

Hewitt, Carl / Inman, Jeff. 1991. "DAI Betwixt and Between: From 'Intelligent Agents' to Open Systems Science." In: IEEE Transactions on Systems, Man, and Cybernetics. 21(6). November–December 1991. pp. 1409–1419.

Hillman, Daniel C. A. / Willis, Deborah J. / Gunawardena, Charlotte N. 1994. "Learner-Interface Interaction in Distance Education: An Extension of Contemporary Models and Strategies for Practitioners." In: The American Journal of Distance Education. 8(2). 1994. pp. 30–42.

Hilton, James L. / von Hippel, William. 1996. "Stereotypes." In: Annual Review of Psychology. 47. February 1996. pp. 237–271.

Hinckley, Ken / Jacob, Robert J. K. / Ware, Colin. 2004. "Input/Output Devices and Interaction Techniques." In: Tucker. (ed.). 2004. pp. 20-1–20-32.

Hirumi, Atsusi. 2002. "A Framework for Analyzing, Designing, and Sequencing Planned E-Learning Interactions." In: The Quarterly Review of Distance Education. 3(2). Summer 2002. pp. 141–160.

Hix, Deborah / Hartson, H. Rex. 1993. Developing User Interfaces: Ensuring Usability Through Product and Process. New York: John Wiley & Sons.

Höfer, Andreas / Tichy, Walter F. 2006. "Status of Empirical Research in Software Engineering." In: Proceedings of the International Workshop on Empirical Software Engineering Issues: Critical Assessment and Future Directions. Castle Dagstuhl, Wadern, Germany. 26–30 June 2006.

Hoffmann, Donna L. / Novak, Thomas P. / Chatterjee, Patrali. 1995. "Commercial Scenarios for the Web: Opportunities and Challenges." In: Journal of Computer-Mediated Communication. 1(3). December 1995. Available Online. URL: http://jcmc.indiana.edu/vol1/issue3/hoffman.html. Last Access: 8 September 2009.

Holmes, John N. / Holmes, Wendy J. 2001. Speech Synthesis and Recognition (2nd ed.). London: Taylor & Francis.

Holt, Peter / Dubs, Shelli / Jones, Marlene / Greer, Jim E. 1994. "The State of Student Modelling." In: Greer & McCalla. (eds.). 1994. pp. 3–35.

Horn, Robert E. / Yoshimi, Jeffrey / Deering, Mark / McBride, Russell. 1998. Mapping Great Debates: Can Computers Think?. Bainbridge Island, Washington: MacroVU Press.

Horrocks, Ian. 2002. "An Ontology Language for the Semantic Web." In: IEEE Intelligent Systems. 17(2). March–April 2002. pp. 74–75.

Horvitz, Eric J. / Breese, John S. / Henrion, Max. 1988. "Decision Theory in Expert Systems and Artificial Intelligence." Technical Report KSL-88-13. Knowledge Systems Laboratory, Stanford University. Stanford, California, USA. July 1988. Available Online. URL: ftp://ftp.research.microsoft .com/pub/ejh/dt.pdf. Last Access: 8 September 2009.

Horvitz, Eric J. / Breese, Jack / Heckerman, David / Hovel, David / Rommelse, Koos. 1998. "The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users." In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI '98). Madison, Wisconsin, USA. 24–26 July 1998.

Houston, Archester. 1997. "Survey Handbook." TQLO Publication 97-06. Department of the Navy, Total Quality Leadership Office (TQLO). Washington, District of Columbia, USA. Available Online. URL: http://www.uiowa.edu/~cqi/surveybk.pdf. Last Access: 8 September 2009.

Hovy, Eduard H. 1993. "Automated Discourse Generation Using Discourse Structure Relations." In: Artificial Intelligence. 63(1–2). October 1993. pp. 341–385.

Hovy, Eduard H. 1997. "Language Generation: Overview." In: Cole et al. (eds.). 1997. pp. 139–146.

Hovy, Eduard H. 2001. "Natural Language Generation." In: Wilson & Keil. (eds.). 2001. pp. 589–592.

Hovy, Eduard H. 2003. "Text Summarization." In: Mitkov. (ed.). 2003. pp. 583–598.

Hovy, Eduard H. / Wanner, Leo. 1996. "Managing Sentence Planning Requirements." In: Proceedings of the Workshop "Gaps and Bridges: New Directions in Planning and Natural Language Generation." 12th European Conference on Artificial Intelligence (ECAI '96). Budapest, Hungary. 11–16 August 1996.

Huang, Amy / Lee, Francis / Nass, Clifford I. / Paik, Young / Swartz, Luke. 2000. "Can Voice User Interfaces Say "I"? – An Experiment with Recorded Speech and TTS." Department of Communication, Stanford University. Stanford, California, USA. Available Online. URL: http://xenon.stanford.edu/ ~lswartz/IMe_paper.pdf. Last Access: 8 September 2009.

Huang, Thomas S. / Nijholt, Anton / Pantic, Maja / Pentland, Alex P. (eds.). 2007. Artificial Intelligence for Human Computing. Berlin: Springer Verlag.

Huang, Xuedong / Acero, Alex / Hon, Hsiao-Wuen. 2001. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. Upper Saddle River, New Jersey: Prentice Hall.

Hudlicka, Eva. 2003. “To Feel or Not to Feel: The Role of Affect in Human-Computer Interaction.” In: International Journal of Human-Computer Studies. 59(1–2). July 2003. pp. 1–32.

Huhns, Michael N. 1987. Distributed Artificial Intelligence. Los Altos, California: Morgan Kaufmann Publishers.

Hutchby, Ian / Wooffitt, Robin. 1998. Conversation Analysis. Cambridge: Polity Press.

Hutchins, W. John / Somers, Harold L. 1992. An Introduction to Machine Translation. London: Academic Press.

Iacobelli, Francisco / Cassell, Justine. 2007. “Ethnic Identity and Engagement in Embodied Conversational Agents.” In: Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA 2007). Paris, France. 17–19 September 2007.

Isard, Stephen. 2001. “Speech Synthesis.” In: Wilson & Keil. (eds.). 2001. pp. 792–794.

Isbister, Katherine. 1995. “Perceived Intelligence and the Design of Computer Characters.” M.A. Thesis. Department of Communication, Stanford University. Stanford, California, USA. 1995. Presented at the Workshop on Lifelike Computer Characters (LCC ’95). Snowbird, Utah, USA. 26–29 September 1995. No Longer Available Online. URL: http://www.katherineinterface.com/ lccpaper.html. Last Access: 7 April 2007.

Isbister, Katherine. 2006. Better Game Characters by Design: A Psychological Approach. San Francisco: Morgan Kaufmann Publishers.

Isbister, Katherine / Doyle, Patrick O. 1999. “Touring Machines: Guide Agents for Sharing Stories About Digital Places.” In: Proceedings of the AAAI Fall Symposium on Narrative Intelligence. North Falmouth, Massachusetts, USA. 5–7 November 1999.

Isbister, Katherine / Doyle, Patrick O. 2002. “Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy.” In: Proceedings of the Workshop “Embodied Conversational Agents – Let’s Specify and Evaluate Them!” 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Isbister, Katherine / Doyle, Patrick O. 2004. “The Blind Men and the Elephant Revisited: Evaluating Interdisciplinary ECA Research.” In: Ruttkay & Pelachaud. (eds.). 2004. pp. 3–26.

Isbister, Katherine / Hayes-Roth, Barbara. 1998. “Social Implications of Using Synthetic Characters: An Examination of a Role-Specific Intelligent Agent.” Technical Report KSL-98-01. Stanford Knowledge Systems Laboratory, Stanford University. Stanford, California, USA. January 1998. Available Online. URL: ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-98-01.ps.gz. Last Access: 8 September 2009.

Isbister, Katherine / Nass, Clifford I. 2000. “Consistency of Personality in Interactive Characters: Verbal Cues, Non-Verbal Cues, and User Characteristics.” In: International Journal of Human-Computer Studies. 53(2). August 2000. pp. 251–267.

Isen, Alice M. 2000. “Positive Affect and Decision Making.” In: Lewis & Haviland-Jones. (eds.). 2000. pp. 417–435.

Isen, Alice M. 2002. “A Role for Neuropsychology in Understanding the Facilitating Influence of Positive Affect on Social Behavior and Cognitive Processes.” In: Snyder & Lopez. (eds.). 2002. pp. 528–540.

Ishiguro, Hiroshi. 2005. “Android Science: Toward a New Cross-Interdisciplinary Framework.” In: Proceedings of the Cognitive Science Society 2005 Workshop “Toward Social Mechanisms of Android Science.” Stresa, Italy. 25–26 July 2005.

Isla, Damián / Blumberg, Bruce. 2002. “New Challenges for Character-Based AI for Games.” In: Proceedings of the AAAI Spring Symposium on AI and Interactive Entertainment. Stanford, California, USA. 25–27 March 2002.

ISO. 1998. “Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs), Part 11: Guidance on Usability.” ISO 9241-11:1998. International Organization for Standardization (ISO). Geneva, Switzerland.

Issing, Ludwig J. 1995. “Instruktionsdesign für Multimedia.” In: Issing & Klimsa. (eds.). 1995. pp. 195–220.

Issing, Ludwig J. / Klimsa, Paul. (eds.). 1995. Information und Lernen mit Multimedia. Weinheim, Germany: Psychologie Verlags Union.

Jacko, Julie A. / Sears, Andrew. (eds.). 2003. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Jackson, Peter. 1999. Introduction to Expert Systems (3rd ed.). Harlow, United Kingdom: Addison-Wesley Longman.

Jaimes, Alejandro / Sebe, Nicu. 2005. “Multimodal Human-Computer Interaction: A Survey.” In: Proceedings of the IEEE International Workshop on Human-Computer Interaction. Beijing, China. 21 October 2005.

James, William. 1890. The Principles of Psychology. 2 Volumes. New York: Henry Holt.

Jameson, Anthony. 1995. “Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues.” In: User Modeling and User-Adapted Interaction. 5(3–4). September 1995. pp. 193–251.

Jameson, Anthony. 2003. “Adaptive Interfaces and Agents.” In: Jacko & Sears. (eds.). 2003. pp. 305–330.

Jaspers, Fons. 1991. “Interactivity or Instruction?: A Reaction to Merrill.” In: Educational Technology. 31(3). March 1991. pp. 21–24.

Jelinek, Frederick. 1998. Statistical Methods for Speech Recognition. Cambridge, Massachusetts: The MIT Press.

Jennings, Nicholas R. / Sycara, Katia P. / Wooldridge, Michael J. 1998. “A Roadmap of Agent Research and Development.” In: Autonomous Agents and Multi-Agent Systems. 1(1). March 1998. pp. 7–38.

Jennings, Nicholas R. / Wooldridge, Michael J. 1996. “Software Agents.” In: IEE Review. 42(1). January 1996. pp. 17–20.

Jennings, Nicholas R. / Wooldridge, Michael J. (eds.). 1998a. Agent Technology: Foundations, Applications, and Markets. Berlin: Springer Verlag.

Jennings, Nicholas R. / Wooldridge, Michael J. 1998b. “Applications of Intelligent Agents.” In: Jennings & Wooldridge. (eds.). 1998a. pp. 3–27.

Jensen, Finn V. 1996. An Introduction to Bayesian Networks. New York: Springer Verlag.

Jensen, Jens F. 1998. “Interactivity: Tracking a New Concept in Media and Communication Studies.” In: Nordicom Review. 19(1). 1998. pp. 185–204.

John, Oliver P. / Srivastava, Sanjay. 1999. “The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives.” In: Pervin & John. (eds.). 1999. pp. 102–138.

Johnson, Addie / Taatgen, Niels. 2005. “User Modeling.” In: Proctor & Vu. (eds.). 2005. pp. 424–437.

Johnson, Gerald J. 1994. “Of Metaphor and the Difficulty of Computer Discourse.” In: Communications of the ACM. 37(12). December 1994. pp. 97–102.

Johnson, Richard D. / Marakas, George M. / Palmer, Jonathan W. 2006. “Differential Social Attributions Toward Computing Technology: An Empirical Investigation.” In: International Journal of Human-Computer Studies. 64(5). May 2006. pp. 446–460.

Johnson, W. Lewis. 1998. “Pedagogical Agents.” In: AI*IA Notizie. 11(3). September 1998. pp. 50–57.

Johnson, W. Lewis. 2003. “Using Agent Technology to Improve the Quality of Web-Based Education.” In: Zhong et al. (eds.). 2003. pp. 77–103.

Johnson, W. Lewis. 2004. “System-Level Issues in Evaluation of ECAs.” Presentation. Working Group 4. Dagstuhl Seminar 04121 on Evaluating Embodied Conversational Agents. Castle Dagstuhl, Wadern, Germany. 15–19 March 2004.

Johnson, W. Lewis. 2007. “Serious Use of a Serious Game for Language Learning.” In: Proceedings of the 13th International Conference on Artificial Intelligence in Education (AI-ED 2007). Marina del Rey, California, USA. 9–13 July 2007.

Johnson, W. Lewis / Barker, Tim / Bernsen, Niels O. / Biswas, Gautam / Cavazza, Marc / Christoph, Noor / Gerhard, Michael / Nijholt, Anton / Prendinger, Helmut. 2004a. “System-Level Evaluation of ECAs.” Final Presentation. Working Group 4. Dagstuhl Seminar 04121 on Evaluating Embodied Conversational Agents. Castle Dagstuhl, Wadern, Germany. 15–19 March 2004. Available Online. URL: http://wwwhome.cs.utwente.nl/~zsofi/eeca/WG4.ppt. Last Access: 8 September 2009.

Johnson, W. Lewis / Beal, Carole / Fowles-Winkler, Anna / Lauper, Ursula / Marsella, Stacy / Narayanan, Shrikanth / Papchristou, Dimitra / Vilhjálmsson, Hannes H. 2004b. “Tactical Language Training System: An Interim Report.” In: Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS 2004). Maceió, Alagoas, Brazil. 30 August–3 September 2004.

Johnson, W. Lewis / LaBore, Catherine / Chiu, Yuan-Chun. 2004c. “A Pedagogical Agent for Psychosocial Intervention on a Handheld Computer.” In: Proceedings of the AAAI Fall Symposium on Dialogue Systems for Health Communication. Washington, District of Columbia, USA. 21–24 October 2004.

Johnson, W. Lewis / Rizzo, Paola / Bosma, Wauter / Kole, Sander / Ghijsen, Mattijs / van Welbergen, Herwin. 2004d. “Generating Socially Appropriate Tutorial Dialog.” In: Proceedings of the Tutorial and Research Workshop on Affective Dialogue Systems (ADS 2004). Kloster Irsee, Germany. 14–16 June 2004.

Johnson, W. Lewis / Narayanan, Shrikanth / Whitney, Richard / Das, Rajat / Bulut, Murtaza / LaBore, Catherine. 2002. “Limited Domain Synthesis of Expressive Military Speech for Animated Characters.” In: Proceedings of the IEEE Speech Synthesis Workshop. Santa Monica, California, USA. 11–13 September 2002.

Johnson, W. Lewis / Rickel, Jeff W. / Stiles, Randy / Munro, Allen. 1998. “Integrating Pedagogical Agents into Virtual Environments.” In: Presence: Teleoperators and Virtual Environments. 7(6). December 1998. pp. 523–546.

Johnson, W. Lewis / Rickel, Jeff W. / Lester, James C. 2000. “Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments.” In: International Journal of Artificial Intelligence in Education. 11(1). 2000. pp. 47–78.

Johnson, W. Lewis / Rickel, Jeff W. / Lester, James C. 2001. “Research in Animated Pedagogical Agents: Progress and Prospects for Training.” Document Based on a Report Prepared for Naval Air Systems Command, Naval Air Warfare Center Training Systems Division. Orlando, Florida, USA. May 2001. Available Online. URL: ftp://ftp.isi.edu/isd/johnson/agents_review_may_01.doc. Last Access: 8 September 2009.

Johnson, W. Lewis / Shaw, Erin / Marshall, Andrew / LaBore, Catherine. 2003a. “Evolution of User Interaction: The Case of Agent Adele.” In: Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI 2003). Miami, Florida, USA. 12–15 January 2003.

Johnson, W. Lewis / Kole, Sander / Shaw, Erin / Pain, Helen G. 2003b. “Socially Intelligent Learner-Agent Interaction Tactics.” In: Proceedings of the 11th International Conference on Artificial Intelligence in Education (AI-ED 2003). Sydney, Australia. 20–24 July 2003.

Johnson, W. Lewis / Valente, Andre. 2008. “Tactical Language and Culture Training System: Using Artificial Intelligence to Teach Foreign Language and Cultures.” In: Proceedings of the 20th Conference on Innovative Applications of Artificial Intelligence (IAAI 2008). Chicago, Illinois, USA. 13–17 July 2008.

Johnson, W. Lewis / Valente, Andre / Heuts, Robert. 2008. “Multi-Platform Delivery of Game-Based Learning Content.” In: Proceedings of the Society for Applied Learning Technology (SALT) New Learning Technologies Conference. Orlando, Florida, USA. 20–22 February 2008.

Johnson, W. Lewis / Vilhjálmsson, Hannes H. / Marsella, Stacy. 2005a. “Serious Games for Language Learning: How Much Game, How Much AI?.” In: Proceedings of the 12th International Conference on Artificial Intelligence in Education (AI-ED 2005). Amsterdam, The Netherlands. 18–22 July 2005.

Johnson, W. Lewis / Vilhjálmsson, Hannes H. / Samtani, Prasan. 2005b. “The Tactical Language Training System (Technical Demonstration).” In: Proceedings of the 1st Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2005). Marina del Rey, California, USA. 1–3 June 2005.

Jonassen, David H. 1988. Instructional Design for Microcomputer Courseware. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Jonassen, David H. 1991. “Evaluating Constructivistic Learning.” In: Educational Technology. 31(9). September 1991. pp. 28–33.

Jonassen, David H. 1994. “Thinking Technology: Toward a Constructivist Design Model.” In: Educational Technology. 34(4). April 1994. pp. 34–37.

Jonassen, David H. (ed.). 1996. Handbook of Research on Educational Communications and Technology. New York: MacMillan Publishers.

Jonassen, David H. (ed.). 2004. Handbook of Research on Educational Communications and Technology (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Jonassen, David H. / Carr, Chad / Yueh, Hsiu-Ping. 1998. “Computers as Mindtools for Engaging Learners in Critical Thinking.” In: TechTrends. 43(2). March 1998. pp. 24–32.

Jonassen, David H. / Reeves, Thomas C. 1996. “Learning with Technology: Using Computers as Cognitive Tools.” In: Jonassen. (ed.). 1996. pp. 693–719.

Jones, Rhys J. 2003. “First Investigations and Experiments in Welsh Speech Recognition.” Ph.D. Thesis. School of Engineering, University of Wales. Swansea, United Kingdom. March 2003. No Longer Available Online. URL: http://www.bangor.ac.uk/~cbs204/cyhoeddiadau/Jones_2003_PhD.pdf. Last Access: 27 November 2007.

Jones, Warren H. / Hansson, Robert O. / Phillips, Adam L. 1978. “Physical Attractiveness and Judgments of Psychopathology.” In: Journal of Social Psychology. 105(1st Half). June 1978. pp. 79–84.

Jonsson, Ing-Marie / Nass, Clifford I. / Endo, Jack / Reaves, Ben / Harris, Helen / Le Ta, Janice / Chan, Nicolas / Knapp, Sean. 2004. “Don’t Blame Me, I’m Only the Driver: Impact of Blame Attribution on Attitudes and Attention to Driving Task.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2004). Vienna, Austria. 24–29 April 2004.

Juler, Philip. 1990. “Promoting Interaction; Maintaining Independence: Swallowing the Mixture.” In: Open Learning. 5(2). June 1990. pp. 24–33.

Jurafsky, Daniel / Martin, James H. 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, New Jersey: Prentice Hall.

Jurafsky, Daniel / Martin, James H. 2008. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Juslin, Patrik N. / Laukka, Petri. 2003. “Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?.” In: Psychological Bulletin. 129(5). September 2003. pp. 770–814.

Kabassi, Katerina / Virvou, Maria. 2003. “Using Web Services for Personalised Web-Based Learning.” In: Educational Technology & Society. 6(3). July 2003. pp. 61–71.

Kaplan, Ronald M. 2003. “Syntax.” In: Mitkov. (ed.). 2003. pp. 70–90.

Kappas, Arvid / Hess, Ursula / Scherer, Klaus R. 1991. “Voice and Emotion.” In: Feldman & Rimé. (eds.). 1991. pp. 200–238.

Karlsen, Jan E. / Karlsen, Hanne. 2007. “Expert Groups as Production Units for Shared Knowledge in Energy Foresights.” In: foresight. 9(1). 2007. pp. 37–49.

Karrer, Tony. 2007. “Understanding E-Learning 2.0.” In: Learning Circuits. 7 July 2007. Available Online. URL: http://www.astd.org/LC/2007/0707_karrer.htm. Last Access: 8 September 2009.

Karttunen, Lauri. 1983. “Kimmo: A General Morphological Processor.” In: Texas Linguistics Forum. 22. 1983. pp. 165–186.

Karttunen, Lauri. 2003. “Finite-State Technology.” In: Mitkov. (ed.). 2003. pp. 339–357.

Karttunen, Lauri / Beesley, Kenneth R. 2005. “Twenty-Five Years of Finite-State Morphology.” In: Arppe et al. (eds.). 2005. pp. 71–83.

Katagiri, Yasuhiro / Nass, Clifford I. / Takeuchi, Yugo. 2001. “Cross-Cultural Studies of the Computers Are Social Actors Paradigm: The Case of Reciprocity.” In: Proceedings of the 9th International Conference on Human-Computer Interaction (HCI 2001). New Orleans, Louisiana, USA. 5–10 August 2001.

Katzlberger, Thomas. 2005. “Learning by Teaching Agents.” Ph.D. Thesis. Department of Electrical Engineering and Computer Science, Vanderbilt University. Nashville, Tennessee, USA. December 2005. Available Online. URL: http://www.teachableagents.org/papers/ThomasKatzlbergerDis-1.17-ds.pdf. Last Access: 8 September 2009.

Kay, Alan C. 1984. “Computer Software.” In: Scientific American. 251(3). September 1984. pp. 41–47.

Kay, Alan C. 1990. “User Interface: A Personal View.” In: Laurel. (ed.). 1990a. pp. 191–207.

Kay, Alan C. 1991. “Computers, Networks and Education.” In: Scientific American. 265(3). Special Issue. September 1991. pp. 138–148. Available Online. URL: http://www.squeakland.org/content/articles/attach/sci_amer_article.pdf. Last Access: 8 September 2009.

Kay, Judy. 1997. “Learner Know Thyself: Student Models to Give Learner Control and Responsibility.” Invited Keynote Address at the 5th International Conference on Computers in Education (ICCE ’97). Kuching, Sarawak, Malaysia. 2–6 December 1997.

Kay, Judy. 2000. “Stereotypes, Student Models and Scrutability.” In: Proceedings of the 5th International Conference on Intelligent Tutoring Systems (ITS 2000). Montreal, Quebec, Canada. 19–23 June 2000.

Kay, Martin / Gawron, Jean M. / Norvig, Peter. 1994. Verbmobil: A Translation System for Face-to-Face Dialog. Stanford, California: CSLI Publications.

Kearsley, Greg. 2000. Online Education: Learning and Teaching in Cyberspace. Belmont, California: Wadsworth Publishing.

Keegan, Desmond. (ed.). 1993. Theoretical Principles of Distance Education. New York: Routledge.

Keegan, Desmond. 2002. “The Future of Learning: From eLearning to mLearning.” Technical Report 119 (ZIFF PAPIERE 119). Institute for Research into Distance Education, FernUniversität in Hagen. Hagen, Germany. November 2002. Available Online. URL: http://www.fernuni-hagen.de/ZIFF/ZP_119.pdf. Last Access: 8 September 2009.

Keller, Bill. 1992. “Formalisms for Grammatical Knowledge Representation.” In: Artificial Intelligence Review. 6(4). December 1992. pp. 365–381.

Kelley, John F. 1984. “An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications.” In: ACM Transactions on Information Systems. 2(1). January 1984. pp. 26–41.

Kemshal-Bell, Guy. 2001. “The Online Teacher.” Final Report Prepared for the Project Steering Committee of the VET Teachers and Online Learning Project. New South Wales Department of Education and Training. Sydney, Australia. April 2001. Available Online. URL: http://www.staff.vu.edu.au/ronj/PDforOLTeachers.pdf. Last Access: 8 September 2009.

Kendal, Simon / Creen, Malcolm. 2007. An Introduction to Knowledge Engineering. London: Springer Verlag.

Kendon, Adam. 1995. “Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation.” In: Journal of Pragmatics. 23(3). March 1995. pp. 247–279.

Kenney, Keith / Gorelik, Alexander / Mwangi, Sam. 2000. “Interactive Features of Online Newspapers.” In: First Monday. 5(1). 3 January 2000. Available Online. URL: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/720/629. Last Access: 8 September 2009.

Kerly, Alice / Hall, Phil / Bull, Susan. 2006. “Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models.” In: Proceedings of the 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (AI 2006). Cambridge, United Kingdom. 11–13 December 2006.

Kerres, Michael. 1998. Multimediale und Telemediale Lernumgebungen: Konzeption und Entwicklung. Munich: Oldenbourg Verlag.

Kerres, Michael / de Witt, Claudia. 2003. "A Didactical Framework for the Design of Blended Learning Arrangements." In: Journal of Educational Media. 28(2–3). October 2003. pp. 101–113.

Kim, Anya / Hoffman, Lance J. / Martin, C. Dianne. 2002. "Building Privacy into the Semantic Web: An Ontology Needed Now." In: Proceedings of the International Workshop on the Semantic Web. 11th International World Wide Web Conference (WWW 2002). Honolulu, Hawaii, USA. 7–11 May 2002.

Kim, Yanghee. 2003. "Pedagogical Agent as Learning Companion: Its Constituents and Educational Implications." In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2003). Phoenix, Arizona, USA. 7–11 November 2003.

Kim, Yanghee. 2005. "Empathetic Virtual Peers Enhanced Learner Interest and Self-Efficacy." In: Proceedings of the Workshop on Motivation and Affect in Educational Software. 12th International Conference on Artificial Intelligence in Education (AI-ED 2005). Amsterdam, The Netherlands. 18–22 July 2005.

Kim, Yanghee. 2007. "Desirable Characteristics of Learning Companions." In: International Journal of Artificial Intelligence in Education. 17(4). 2007. pp. 371–388.

Kim, Yanghee / Baylor, Amy L. 2005a. "The Impact of Effective Expression and Gender of a Learning Companion." In: Proceedings of the Annual Conference of the American Educational Research Association (AERA 2005). Montreal, Quebec, Canada. 11–15 April 2005.

Kim, Yanghee / Baylor, Amy L. 2005b. "Pedagogical Agents as Learning Companions: Building Empathetic Relationships with Learners." In: Proceedings of the Annual Conference of the American Educational Research Association (AERA 2005). Montreal, Quebec, Canada. 11–15 April 2005.

Kim, Yanghee / Baylor, Amy L. 2006. "A Social-Cognitive Framework for Pedagogical Agents as Learning Companions." In: Educational Technology Research and Development. 54(6). December 2006. pp. 569–596.

Kim, Yanghee / Baylor, Amy L. / The PALS Group. 2006. "Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction." In: Educational Technology Research and Development. 54(3). June 2006. pp. 223–243.

Kim, Yanghee / Baylor, Amy L. / Shen, E. 2007. "Pedagogical Agents as Learning Companions: The Impact of Agent Emotion and Gender." In: Journal of Computer-Assisted Learning. 23(3). June 2007. pp. 220–234.

Kim, Yanghee / Wei, Quan / Xu, Beijie / Ko, Youngah / Ilieva, Vessela. 2007. "MathGirls: Toward Developing Girls' Positive Attitude and Self-Efficacy Through Pedagogical Agents." In: Proceedings of the 13th International Conference on Artificial Intelligence in Education (AI-ED 2007). Marina del Rey, California, USA. 9–13 July 2007.

King, William J. / Ohya, Jun. 1995. "The Representation of Agents: A Study of Phenomena in Virtual Environments." In: Proceedings of the 4th IEEE International Workshop on Robot and Human Communication (RO-MAN '95). Tokyo, Japan. 5–7 July 1995.

Kinzer, Charles K. / Verhoeven, Ludo. (eds.). 2007. Interactive Literacy Education: Facilitating Literacy Environments Through Technology. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kiousis, Spiro. 2002. "Interactivity: A Concept Exploration." In: New Media & Society. 4(3). September 2002. pp. 355–383.

Kirschner, Paul A. / Kreijns, Karel / Jochems, Wim. 2003. "Measuring the Perceived Quality of Social Space in Distributed Learning Groups." In: Wasson et al. (eds.). 2003. pp. 323–332.

Kleinedler, Steven. (ed.). 2001. Dictionary of Computer and Internet Words: An A to Z Guide to Hardware, Software, and Cyberspace. Boston, Massachusetts: Houghton Mifflin Company.

Kleinginna, Paul R., Jr. / Kleinginna, Anne M. 1981. "A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition." In: Motivation and Emotion. 5(4). December 1981. pp. 345–379.

Kleinke, Chris L. 1986. "Gaze and Eye Contact: A Research Review." In: Psychological Bulletin. 100(1). July 1986. pp. 78–100.

Knapp, Mark L. / Hall, Judith A. 2005. Nonverbal Communication in Human Interaction (6th ed.). Belmont, California: Wadsworth Publishing.

Knight, Kevin. 1989. "Unification: A Multidisciplinary Survey." In: ACM Computing Surveys. 21(1). March 1989. pp. 93–124.

Ko, Susan / Rossen, Steve. 2001. Teaching Online: A Practical Guide. New York: Hoghton Mifflin.

Kobsa, Alfred / Wahlster, Wolfgang. (eds.). 1989. User Models in Dialog Systems. Berlin: Springer Verlag.

Kobsa, Alfred. 2001. "Generic User Modeling Systems." In: User Modeling and User-Adapted Interaction. 11(1–2). March 2001. pp. 49–63.

Koda, Tomoko / Maes, Pattie. 1996. "Agents with Faces: The Effects of Personification of Agents." In: Proceedings of the 11th British Computer Society Conference on Human-Computer Interaction (HCI '96). London, United Kingdom. 20–23 August 1996.

Koedinger, Kenneth R. 2001. "Cognitive Tutors as Modeling Tools and Instructional Models." In: Forbus & Feltovich. (eds.). 2001. pp. 145–167.

Koedinger, Kenneth R. / Anderson, John R. / Hadley, William H. / Mark, Mary A. 1997. "Intelligent Tutoring Goes to School in the Big City." In: International Journal of Artificial Intelligence in Education. 8(1). 1997. pp. 30–43.

Koolstra, Cees M. / Bos, Mark J. 2006. "The Development of a New Objective Instrument to Measure Different Levels of Interactivity." In: Proceedings of the 9th International Conference on Public Communication of Science and Technology (PCST-9). Seoul, Korea. 17–20 May 2006.

Koper, Robert. 2004. "Use of the Semantic Web to Solve Some Basic Problems in Education: Increase Flexible, Distributed Lifelong Learning, Decrease Teachers' Workload." In: Journal of Interactive Media in Education. Special Issue on the Educational Semantic Web. 2004(6). 21 May 2004. Available Online. URL: http://www-jime.open.ac.uk/2004/6/koper-2004-6.pdf. Last Access: 8 September 2009.

Kort, Barry / Reilly, Robert / Picard, Rosalind W. 2001a. "An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy – Building a Learning Companion." In: Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2001). Madison, Wisconsin, USA. 6–8 August 2001.

Kort, Barry / Reilly, Robert / Picard, Rosalind W. 2001b. "External Representation of Learning Process and Domain Knowledge: Affective State as a Determinate of Its Structure and Function." In: Proceedings of the Workshop "External Representations in AIED: Multiple Forms and Multiple Roles." 10th International Conference on Artificial Intelligence in Education (AI-ED 2001). San Antonio, Texas, USA. 19–23 May 2001.

Korzeniowski, Paul. 2008. "An Emotional Mess." In: Speech Technology Magazine. 1 October 2008. Available Online. URL: http://www.speechtechmag.com/Articles/Editorial/Cover-Story/An-Emotional-Mess-51042.aspx. Last Access: 8 September 2009.

Koskenniemi, Kimmo. 1983. “Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production.” Ph.D. Thesis. Department of General Linguistics, University of Helsinki. Helsinki, Finland. Technical Report 11. December 1983. Available Online. URL: http://www.ling.helsinki.fi/~koskenni/doc/Two-LevelMorphology.pdf. Last Access: 8 September 2009.

Kotz, David / Gray, Robert S. 1999. “Mobile Agents and the Future of the Internet.” In: ACM Operating Systems Review. 33(3). August 1999. pp. 7–13.

Krahmer, Emiel / Swerts, Marc. 2004. “More About Brows: A Cross-Linguistic Study via Analysis-by-Synthesis.” In: Ruttkay & Pelachaud. (eds.). 2004. pp. 191–216.

Krämer, Nicole C. / Bente, Gary. 2001. “Mehr als Usability: (Sozial-) psychologische Aspekte bei der Evaluation von Anthropomorphen Interface-Agenten.” In: i-com: Zeitschrift für Interaktive und Kooperative Medien. 1(1). 2001. pp. 26–31.

Krämer, Nicole C. / Tietz, Bernd / Bente, Gary. 2003. “Effects of Embodied Interface Agents and Their Gestural Activity.” In: Proceedings of the 4th International Working Conference on Intelligent Virtual Agents (IVA 2003). Kloster Irsee, Germany. 15–17 September 2003.

Kreijns, Karel / Kirschner, Paul A. / Jochems, Wim. 2003. “Identifying the Pitfalls for Social Interaction in Computer-Supported Collaborative Learning Environments: A Review of the Research.” In: Computers in Human Behavior. 19(3). May 2003. pp. 335–353.

Kress, Gunther / van Leeuwen, Theo. 2001. Multimodal Discourse: The Modes and Media of Contemporary Communication. London: Edward Arnold.

Kroeber, Alfred L. / Kluckhohn, Clyde. 1952. Culture: A Critical Review of Concepts and Definitions. Cambridge, Massachusetts: Peabody Museum.

Kroetz, Arthur W. 1999. “The Role of Intelligent Agency in Synthetic Instructor and Human Student Dialogue.” Ph.D. Thesis. University of Southern California. Los Angeles, California, USA. Publication Number AI9987579. 1999.

Kröger, Bernd J. 1996. “Von Sprechbewegungen zum Akustischen Signal: Artikulatorische Sprachsynthese.” In: Spektrum der Wissenschaft. 1996(12). December 1996. pp. 105–107.

Krosnick, Jon A. 1999. “Survey Research.” In: Annual Review of Psychology. 50. February 1999. pp. 537–567.

Kuhlen, Rainer. 1991. Hypertext: Ein Nicht-Lineares Medium zwischen Buch und Wissensbank. Berlin: Springer Verlag.

Kurzweil, Raymond. 1990. The Age of Intelligent Machines. Cambridge, Massachusetts: The MIT Press.

Lachman, Richard W. 1997. “Animist Interface: Experiments in Mapping Character Animation to Computer Interface.” M.Sc. Thesis. Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. Available Online. URL: http://ic.media.mit.edu/Publications/Thesis/richlachMS/ PDF/richMS.pdf. Last Access: 8 September 2009.

Ladefoged, Peter. 2001. A Course in Phonetics (4th ed.). New York: Harcourt College Publishers.

Laird, John E. 2001. “Using a Computer Game to Develop Advanced AI.” In: Computer. 34(7). July 2001. pp. 70–75.

Laird, John E. / van Lent, Michael. 2000. “The Role of AI in Computer Game Genres.” Artificial Intelligence Laboratory, Electrical Engineering and Computer Science Department, University of

Michigan. Ann Arbor, Michigan, USA. Available Online. URL: http://ai.eecs.umich.edu/people/laird/papers/book-chapter.htm. Last Access: 8 September 2009.

Laird, John E. / van Lent, Michael. 2001. "Human-Level AI's Killer Application: Interactive Computer Games." In: AI Magazine. 22(2). Summer 2001. pp. 15–26.

Lajoie, Susanne P. (ed.). 2000. Computers as Cognitive Tools, Volume 2: No More Walls. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lajoie, Susanne P. / Derry, Sharon J. (eds.). 1993. Computers as Cognitive Tools. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lakemeyer, Gerhard / Nebel, Bernhard. (eds.). 2003. Exploring Artificial Intelligence in the New Millennium. San Francisco: Morgan Kaufmann Publishers.

Lakoff, George / Johnson, Mark. 1980. Metaphors We Live By. Chicago: University of Chicago Press.

Lamel, Lori / Gauvain, Jean-Luc. 2003. "Speech Recognition." In: Mitkov. (ed.). 2003. pp. 305–322.

Landow, George P. 2006. Hypertext 3.0: Critical Theory and New Media in an Era of Globalization (3rd ed.). Baltimore, Maryland: John Hopkins University Press.

Lang, Peter J. 1995. "The Emotion Probe: Studies of Motivation and Attention." In: American Psychologist. 50(5). May 1995. pp. 372–385.

Lange, Danny B. / Oshima, Mitsuru. 1998. "Introduction to Mobile Agents." In: Personal and Ubiquitous Computing. 2(2). June 1998. pp. 49–56.

Lange, Danny B. / Oshima, Mitsuru. 1999. "Seven Good Reasons for Mobile Agents." In: Communications of the ACM. 42(3). March 1999. pp. 88–89.

Langer, Ellen J. 1989. Mindfulness. Reading, Massachusetts: Addison-Wesley.

Langer, Ellen J. 1992. "Matters of Mind: Mindfulness / Mindlessness in Perspective." In: Consciousness and Cognition. 1(3). September 1992. pp. 289–305.

Langley, Pat. 1996. Elements of Machine Learning. San Francisco: Morgan Kaufmann Publishers.

Langlois, Judith H. / Kalakanis, Lisa / Rubenstein, Adam J. / Larson, Andrea / Hallam, Monica / Smoot, Monica. 2000. "Maxims or Myths of Beauty?: A Meta-Analytic and Theoretical Review." In: Psychological Bulletin. 126(3). May 2000. pp. 390–423.

Langton, Christopher G. (ed.). 1995, 1997. Artificial Life: An Overview. 1st Paperback Edition. Cambridge, Massachusetts: The MIT Press.

Lanier, Jaron. 1995. "Agents of Alienation." In: Interactions. 2(3). July 1995. pp. 66–72.

Laskey, Kathryn B. 2008. "Lecture Notes on Computational Models for Probabilistic Inference, Unit 4: Belief Propagation in Bayesian Networks: Propagation in Junction Trees." Spring 2008. George Mason University. Fairfax, Virginia, USA. Available Online. URL: http://volgenau.gmu.edu/~klaskey/CompProb/CompProb_JTProp.pdf. Last Access: 8 September 2009.

Lassila, Ora / Hendler, James. 2007. "Embracing 'Web 3.0'." In: IEEE Internet Computing. 11(3). May–June 2007. pp. 90–93.

Laufer, László / Tatai, Gábor. 2004. "Learn, Chat and Play – An ECA Supported Stock Markets E-Learning Curricula." In: Proceedings of the International Conference on Web-Based Education (WBE 2004). International Association of Science and Technology for Development (IASTED). Innsbruck, Austria. 16–18 February 2004.

Laurel, Brenda. (ed.). 1990a. The Art of Human-Computer Interface Design. Reading, Massachusetts: Addison-Wesley.

Laurel, Brenda. 1990b. “Interface Agents: Metaphors with Character.” In: Laurel. 1990a. pp. 355–365.

Laurel, Brenda. 1991. Computers as Theatre. Reading, Massachusetts: Addison-Wesley.

Laurel, Brenda. 1992. “In Defense of Anthropomorphism.” Panel “Anthropomorphism: From Eliza to Terminator 2.” Conference on Human Factors in Computing Systems (CHI ’92). Monterey, California, USA. 3–7 May 1992.

Lauría, Eitel J.M. / Duchessi, Peter J. 2006. “A Bayesian Belief Network for IT Implementation Decision Support.” In: Decision Support Systems. 42(3). December 2006. pp. 1573–1588.

Lave, Jean / Wenger, Etienne. 1991. Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press.

Lazarus, Richard S. 1991. Emotion and Adaptation. New York: Oxford University Press.

Lea, Martin. (ed.). 1992. Contexts of Computer-Mediated Communication. New York: Harvester Wheatsheaf.

Leary, Timothy. 1990. “The Interpersonal, Interactive, Interdimensional Interface.” In: Laurel. (ed.). 1990a. pp. 229–234.

Lee, Eun-Ju / Nass, Clifford I. 1998. “Does the Ethnicity of a Computer Agent Matter?: An Experimental Comparison of Human-Computer Interaction and Computer-Mediated Communication.” In: Proceedings of the 1st Workshop on Embodied Conversational Characters (WECC 1998). Lake Tahoe, California, USA. 12–15 October 1998.

Lee, Jae-Shin. 2000. “Interactivity: A New Approach.” In: Proceedings of the Annual Conference of the Association for Education in Journalism and Mass Communication (AEJMC 2000). Phoenix, Arizona, USA. 9–12 August 2000.

Lee, Jong-Eun R. / Nass, Clifford I. / Brave, Scott / Morishima, Yasunori / Nakajima, Hiroshi / Yamada, Ryota. 2007a. “The Case for Caring Colearners: The Effects of a Computer-Mediated Colearner Agent on Trust and Learning.” In: Journal of Communication. 57(2). June 2007. pp. 183–204.

Lee, Sunyoung / McQuiggan, Scott W. / Lester, James C. 2007b. “Inducing User Affect Recognition Models for Task-Oriented Environments.” In: Proceedings of the 11th International Conference on User Modeling (UM 2007). Corfu, Greece. 25–29 July 2007.

Lee, Kwan Min. 2004. “Presence, Explicated.” In: Communication Theory. 14(1). February 2004. pp. 27–50.

Lee, Kwan Min / Nass, Clifford I. 2004. “The Multiple Source Effect and Synthesized Speech: Doubly-Disembodied Language as a Conceptual Framework.” In: Human Communication Research. 30(2). April 2004. pp. 182–207.

Lee, Kwan Min / Nass, Clifford I. 2005. “Social-Psychological Origins of Feelings of Presence: Creating Social Presence with Machine-Generated Voices.” In: Media Psychology. 7(1). February 2005. pp. 31–45.

Lee, Kwan Min / Jung, Younbo / Kim, Jaywoo / Kim, Sang Ryong. 2006. “Are Physically Embodied Social Agents Better than Disembodied Social Agents?: The Effects of Physical Embodiment, Tactile Interaction, and People’s Loneliness in Human-Robot Interaction.” In: International Journal of Human-Computer Studies. 64(10). October 2006. pp. 962–973.

Lee, Sooha P. / Badler, Jeremy B. / Badler, Norman I. 2002. "Eyes Alive." In: ACM Transactions on Graphics. 21(3). July 2002. pp. 637–644.

Leech, Geoffrey N. 1983. Principles of Pragmatics. London: Longman.

Leech, Geoffrey N. / Weisser, Martin. 2003. "Pragmatics and Dialogue." In: Mitkov. (ed.). 2003. pp. 136–156.

Lehnert, Wendy G. / Ringle, Martin H. (eds.). 1982. Strategies for Natural Language Processing. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Leone, Jan / Sutarso, Pudjiati / Badley, Bill / Whitaker, Donald / Ryan, Linda. 2006. "Institutional Effectiveness Handbook: Assessment in Action: A Guide Toward Institutional Improvement." Office of Institutional Effectiveness, Planning, and Research. Middle Tennessee State University. Murfreesboro, Tennessee, USA. January 2006. Available Online. URL: http://www.mtsu.edu/~iepr/handbook.pdf. Last Access: 8 September 2009.

Lepper, Mark R. / Henderlong, Jennifer. 2000. "Turning 'Play' into 'Work' and 'Work' into 'Play:' 25 Years of Research on Intrinsic Versus Extrinsic Motivation." In: Sansone & Harackiewicz (eds.). 2000. pp. 257–307.

Lester, James C. / Callaway, Charles B. / Grégoire, Joël P. / Stelling, Gary D. / Towns, Stuart C. Zettlemoyer, Luke S. 2001. "Animated Pedagogical Agents in Knowledge-Based Learning Environments." In: Forbus & Feltovich. (eds.). 2001. pp. 269–298.

Lester, James C. / Callaway, Charles B. / Stone, Brian A. / Towns, Stuart G. 1997a. "Mixed Initiative Problem Solving with Animated Pedagogical Agents." In: Working Notes of the AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction. Stanford, California, USA. 24–26 March 1997.

Lester, James C. / Converse, Sharolyn A. / Kahler, Susan E. / Barlow, S. Todd / Stone, Brian A. / Bhogal, Ravinder S. 1997b. "The Persona Effect: Affective Impact of Animated Pedagogical Agents." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI '97). Atlanta, Georgia, USA. 22–27 March 1997.

Lester, James C. / Converse, Sharolyn A. / Stone, Brian A. / Kahler, Susan E. / Barlow, S. Todd. 1997c. "Animated Pedagogical Agents and Problem-Solving Effectiveness: A Large-Scale Empirical Evaluation." In: Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED '97). Kobe, Japan. 18–22 August 1997.

Lester, James C. / Stone, Brian A. / Stelling, Gary D. 1999a. "Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments." In: User Modeling and User-Adapted Interaction. 9(1–2). April 1999. pp. 1–44.

Lester, James C. / Voerman, Jennifer L. / Towns, Stuart G. / Callaway, Charles B. 1999b. "Deictic Believability: Coordinated Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents." In: Applied Artificial Intelligence. 13(4–5). May 1999. pp. 383–414.

Lester, James C. / Towns, Stuart G. / Callaway Charles B. / Voerman, Jennifer L. / FitzGerald, Patrick J. 2000. "Deictic and Emotive Communication in Animated Pedagogical Agents." In: Cassell et al. (eds.). 2000a. pp. 123–154.

Lester, James C. / Branting, Karl / Bradford, Mott. 2004. "Conversational Agents." In: Singh. (ed.). 2004. pp. 10-1–10-16.

Levelt, Willem J. M. 1995. Speaking: From Intention to Articulation (4th ed.). Cambridge, Massachusetts: The MIT Press.

Levinson, Stephen C. 1983. Pragmatics. Cambridge: Cambridge University Press.

Lewis, Michael. 1998. "Designing for Human-Agent Interaction." In: AI Magazine. 19(2). Summer 1998. pp. 67–78.

Lewis, Michael / Haviland-Jones, Jeannette M. (eds.). 2000. Handbook of Emotions (2nd ed.). New York: The Guilford Press.

Lewis-Beck, Michael S. / Bryman, Alan / Liao, Tim F. (eds.). 2004. "The Sage Encyclopedia of Social Science Research Methods." 3 Volumes. Thousand Oaks, California: SAGE Publications.

Li, Stan Z. / Jain, Anil K. (eds.). 2005. Handbook of Face Recognition. New York: Springer Verlag.

Liaw, Shu-sheng / Huang, Hsiu-mei. 2000. "Enhancing Interactivity in Web-Based Instruction: A Review of the Literature." In: Educational Technology. 40(3). May–June 2000. pp. 41–45.

Lieberman, Henry. 2001. "Interfaces that Give and Take Advice." In: Carroll. (ed.). 2001. pp. 475–485.

Lievrouw, Leah A. / Livingston, Sonia. (eds.). 2002. The Handbook of New Media. London: SAGE Publications.

Liggett, John. 1974. The Human Face. New York: Stein and Day.

Lisetti, Christine L. 2002. "Personality, Affect and Emotion Taxonomy for Socially Intelligent Agents." In: Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2002). Pensacola Beach, Florida, USA. 14–16 May 2002.

Lisetti, Christine L. / Schiano, Diane J. 2000. "Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence, and Cognitive Science Intersect." In: Pragmatics and Cognition. 8(1). January 2000. pp. 185–235.

Liu, Charlotte H. / Matthews, Robert. 2005. "Vygotsky's Philosophy: Constructivism and Its Criticisms Examined." In: International Education Journal. 6(3). July 2005. pp. 386–399.

Liu, Yuping P. 2003. "Developing a Scale to Measure the Interactivity of Web Sites." In: Journal of Advertising Research. 43(2). June 2003. pp. 207–216.

Liu, Yuping P. / Shrum, L. J. 2002. "What Is Interactivity and Is It Always Such a Good Thing?: Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness." In: Journal of Advertising. 31(4). Winter 2002. pp. 53–64.

Logan, Brian. 2008a. "Reactive Architectures I." G53DIA: Designing Intelligent Agents, Lecture 5. School of Computer Science, University of Nottingham. Nottingham, United Kingdom. Available Online. URL: http://www.cs.nott.ac.uk/~bsl/G53DIA/slides/Reactive-Architectures-I.pdf. Last Access: 8 September 2009.

Logan, Brian. 2008b. "Reactive Architectures II." G53DIA: Designing Intelligent Agents, Lecture 6. School of Computer Science, University of Nottingham. Nottingham, United Kingdom. URL: http://www.cs.nott.ac.uk/~bsl/G53DIA/slides/Reactive-Architectures-II.pdf. Last Access: 8 September 2009.

Logan, Brian. 2008c. "Deliberative Architectures I." G53DIA: Designing Intelligent Agents, Lecture 7. School of Computer Science, University of Nottingham. Nottingham, United Kingdom. Available Online. URL: http://www.cs.nott.ac.uk/~bsl/G53DIA/slides/Deliberative-Architectures-I.pdf. Last Access: 8 September 2009.

Logan, Brian. 2008d. "Deliberative Architectures II." G53DIA: Designing Intelligent Agents, Lecture 8. School of Computer Science, University of Nottingham. Nottingham, United Kingdom. Available

Online. URL: http://www.cs.nott.ac.uk/~bsl/G53DIA/slides/Deliberative-Architectures-II.pdf. Last Access: 8 September 2009.

Logan, Brian. 2008e. "Hybrid Architectures." G53DIA: Designing Intelligent Agents, Lecture 9. School of Computer Science, University of Nottingham. Nottingham, United Kingdom. Available Online. URL: http://www.cs.nott.ac.uk/~bsl/G53DIA/slides/Hybrid-Architectures.pdf. Last Access: 8 September 2009.

Lombard, Matthew / Ditton, Theresa. 1997. "At the Heart of It All: The Concept of Presence." In: Journal of Computer-Mediated Communication. 3(2). September 1997. Available Online. URL: http://jcmc.indiana.edu/vol3/issue2/lombard.html. Last Access: 8 September 2009.

Lombard, Matthew / Reich, Robert D. / Grabe, Maria E. / Bracken, Cheryl C. / Ditton, Theresa B. 2000. "Presence and Television: The Role of Screen Size." In: Human Communication Research. 26(1). January 2000. pp. 75–98.

Lombard, Matthew / Snyder-Duch, Jennifer. 2001. "Interactive Advertising and Presence: A Framework." In: Journal of Interactive Advertising. 1(2). Spring 2001. pp. 1–15. Available Online. URL: http://jiad.org/article13. Last Access: 8 September 2009.

Louwerse, Max M. / Graesser, Arthur C. / Lu, Shulan / Mitchell, Heather H. 2005. "Social Cues in Animated Conversational Agents." In: Applied Cognitive Psychology. 19(6). September 2005. pp. 693–704.

Low, Renae / Sweller, John. 2005. "The Modality Principle in Multimedia Learning." In: Mayer. (ed.). 2005a. pp. 147–157.

Loyall, A. Bryan. 1997. "Believable Agents: Building Interactive Personalities." Ph.D. Thesis. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. Technical Report CMU-CS-97-123. May 1997. Available Online. URL: http://www.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/papers/CMU-CS-97-123.pdf. Last Access: 8 September 2009.

Loyall, A. Bryan / Bates, Joseph. 1997. "Personality-Rich Believable Agents that Use Language." In: Proceedings of the 1st International Conference on Autonomous Agents (Agents '97). Marina del Rey, California, USA. 5–8 February 1997.

Luger, George F. 2005. Artificial Intelligence: Structures and Strategies for Complex Problem Solving (5th ed.). Boston, Massachusetts: Addison-Wesley.

MacDorman, Karl F. 2005. "Androids as an Experimental Apparatus: Why Is There an Uncanny Valley and Can We Exploit It?." In: Proceedings of the Cognitive Science Society 2005 Workshop "Toward Social Mechanisms of Android Science." Stresa, Italy. 25–26 July 2005.

MacManus, Richard / Porter, Joshua. 2005. "Web 2.0 for Designers." In: Digital Web Magazine. 4 May 2005. Available Online. URL: http://www.digital-web.com/articles/web_2_for_designers/. Last Access: 8 September 2009.

Maes, Pattie. 1994. "Agents that Reduce Work and Information Overload." In: Communications of the ACM. 37(7). July 1994. pp. 31–40 and 146.

Maes, Pattie. 1995. "Intelligent Software." In: Scientific American. 273(3). September 1995. pp. 84–86.

Maes, Pattie / Guttman, Robert H. / Moukas, Alexandros G. 1999. "Agents that Buy and Sell." In: Communications of the ACM. 42(3). March 1999. pp. 81–91.

Maestri, George. 1996. Digital Character Animation. Indianapolis, Indiana: New Riders Publishing.

Maestri, George. 1999. Digital Character Animation 2, Volume 1: Essential Techniques. Indianapolis, Indiana: New Riders Publishing.

Maestri, George. 2001. Digital Character Animation 2, Volume 2: Advanced Techniques. Indianapolis, Indiana: New Riders Publishing.

Magnenat-Thalmann, Nadia / Thalmann, Daniel. (eds.). 1991. New Trends in Animation and Visualization. Chichester: John Wiley & Sons.

Magnenat-Thalmann, Nadia / Thalmann, Daniel. (eds.). 2001. Deformable Avatars. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Maldonado, Heidy / Hayes-Roth, Barbara. 2004. "Toward Cross-Cultural Believability in Character Design." In: Payr & Trappl. (eds.). 2004. pp. 143–175.

Mangold, Helmut. 1994. "Das Telephon als Intelligenter Gesprächspartner." In: Spektrum der Wissenschaft. 1994(3). March 1994. pp. 97–99.

Mani, Inderjeet. 2001. Automatic Summarization. Amsterdam: John Benjamins.

Manica, David. 2005. "Grand Designs." In: Auditoria Magazine. News Item. 1 October 2005. No Longer Available Online. URL: http://www.auditoria.tv/index.php?page=news&id=1386. Last Access: 16 May 2006.

Mann, Steve / Niedzviecki, Hal. 2001. Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer. New York: Random House Doubleday.

Mann, William C. / Thompson, Sandra. A. 1988. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." In: Text. 8(3). 1988. pp. 243–281.

Manning, Christopher D. / Schütze, Hinrich. 1999. Foundations of Statistical Natural Language Processing. Cambridge, Massachusetts: The MIT Press.

Marakas, George M. / Johnson, Richard D. / Palmer, Jonathan W. 2000. "A Theoretical Model of Differential Social Attributions Toward Computing Technology: When the Metaphor Becomes the Model." In: International Journal of Human-Computer Studies. 52(4). April 2000. pp. 719–750.

Marcus, Sandra L. 2001. "Knowledge Acquisition." In: Wilson & Keil. (eds.). 2001. pp. 428–430.

Markoff, John. 2006. "Entrepreneurs See a Web Guided by Common Sense." In: The New York Times. 12 November 2006. Available Online. URL: http://www.nytimes.com/2006/11/12/business/12web.html. Last Access: 9 September 2009.

Markus, M. Lynne. 1990. "Toward a 'Critical Mass' Theory of Interactive Media." In: Fulk & Steinfeld. (eds.). 1990. pp. 194–218.

Marsella, Stacy / Gratch, Jonathan / Rickel, Jeff W. 2004. "Expressive Behaviors for Virtual Worlds." In: Prendinger & Ishizuka. (eds.). 2004a. pp. 317–360.

Marshall, Catherine C. / Shipman, Frank M. 2003. "Which Semantic Web?." In: Proceedings of the 14th ACM Conference on Hypertext and Hypermedia (HyperText 2003). Nottingham, United Kingdom. 26–30 August 2003.

Martin, David W. 2007. Doing Psychology Experiments (7th ed.). Belmont, California: Wadsworth Publishing.

Martin, Jean-Claude / Grimard, Sarah / Alexandri, Katerina. 2001. "On the Annotation of Multimodal Behavior and Computation of Cooperation Between Modalities." In: Proceedings of the Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents. 5th International Conference on Autonomous Agents (Agents 2001). Montreal, Quebec, Canada. 28 May–01 June 2001.

Martin, Joel / VanLehn, Kurt. 1995. "Student Assessment Using Bayesian Nets." In: International Journal of Human-Computer Studies. 42(6). June 1995. pp. 575–591.

Martínez-Miranda, Juan / Aldea, Arantza. 2005. "Emotions in Human and Artificial Intelligence." In: Computers in Human Behavior. 21(2). March 2005. pp. 323–341.

Martinho, Carlos / Machado, Isabel / Paiva, Ana. 1999. "A Cognitive Approach to Affective User Modeling." In: Proceedings of the Workshop on Affective Interactions: Towards a New Generation of Interfaces. 3rd i3 Annual Conference (i3AC '99). Siena, Italy. 20–22 October 1999.

Maslow, Abraham H. 1943. "A Theory of Human Motivation." In: Psychological Review. 50(4). July 1943. pp. 370–396.

Maslow, Abraham H. 1970. Motivation and Personality (2nd ed.). New York: Harper & Row.

Mason, Robin / Rennie, Frank. 2006. E-Learning: The Key Concepts. London, New York: Routledge.

Massaro, Dominic W. 1998. Perceiving Talking Faces: From Speech Perception to a Behavioral Principle. Cambridge, Massachusetts: The MIT Press.

Massaro, Dominic W. 2005. "The Psychology and Technology of Talking Heads: Applications in Language Learning." In: Van Kuppevelt et al. (eds.). 2005. pp. 183–214.

Massaro, Dominic W. / Cohen, Michael M. / Beskow, Jonas / Cole, Ronald A. 2000. "Developing and Evaluating Conversational Agents." In: Cassell et al. (eds.). 2000a. pp. 287–318.

Massey, Brian L. / Levy, Mark R. 1999. "Interactivity, Online Journalism, and English-Language Web Newspapers in Asia." In: Journalism & Mass Communication Quarterly. 76(1). Spring 1999. pp. 138–151.

Masterton, Simon. 1998. "Computer Support for Learners Using Intelligent Educational Agents: The Way Forward." In: Proceedings of the 6th International Conference on Computers in Education (ICCE '98). Beijing, China. 14–17 October 1998.

Mateas, Michael. 1997. "An Oz-Centric Review of Interactive Drama and Believable Agents." Technical Report CMU-CS-97-156. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. June 1997. Available Online. URL: http://www.cs.cmu.edu/~michaelm/publications/CMU-CS-97-156.pdf. Last Access: 9 September 2009.

Mateas, Michael. 2003. "Expressive AI: Games and Artificial Intelligence." In: Proceedings of the Digital Games Research Conference (DIGRA 2003). Utrecht, The Netherlands. 4–6 November 2003.

Mateas, Michael / Sengers, Phoebe. 1999. "Narrative Intelligence." In: Proceedings of the AAAI Fall Symposium on Narrative Intelligence. North Falmouth, Massachusetts, USA. 5–7 November 1999.

Mathes, Eugene W. 1975. "The Effects of Physical Attractiveness and Anxiety on Heterosexual Attraction Over a Series of Five Encounters." In: Journal of Marriage and the Family. 37(4). November 1975. pp. 769–773.

Mattern, Friedemann (ed.). 2008. Wie Arbeiten die Suchmaschinen von Morgen?: Informationstechnische, Politische und Ökonomische Perspektiven. Stuttgart: Fraunhofer IRB Verlag.

Mauldin, Michael L. 1994. "Chatterbots, Tinymuds, and the Turing Test: Entering the Loebner Prize Competition." In: Proceedings of the 12th National Conference on Artificial Intelligence (AAAI '94). Seattle, Washington, USA. 31 July–4 August 1994.

Maybury, Mark T. (ed.). 1993. Intelligent Multimedia Interfaces. Menlo Park, California: AAAI / The MIT Press.

Maybury, Mark T. 1995. “Research in Multimedia and Multimodal Parsing and Generation.” In: Artificial Intelligence Review. 9(2–3). June 1995. pp. 103–127.

Maybury, Mark T. 2002. “Language Technology: A Survey of the State of the Art. Language Resources: Multimodal Language Resources.” Technical Paper. The MITRE Corporation. Bedford, Massachusetts, USA. December 2002. Available Online. URL: http://www.mitre.org/work/tech_papers/tech_papers_02/maybury_language/maybury_language_tech.pdf. Last Access: 9 September 2009.

Maybury, Mark T. (ed.). 2004. New Directions in Question Answering. Menlo Park, California: AAAI / The MIT Press.

Mayer, Horst O. / Treichel, Dietmar. (eds.). 2004. Handlungsorientiertes Lernen und eLearning: Grundlagen und Praxisbeispiele. Munich: Oldenbourg Verlag.

Mayer, Richard E. 2002. “A Taxonomy for Computer-Based Assessment of Problem Solving.” In: Computers in Human Behavior.” 18(6). November 2002. pp. 623–632.

Mayer, Richard E. (ed.). 2005a. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press.

Mayer, Richard E. 2005b. “Introduction to Multimedia Learning.” In: Mayer. (ed.). 2005a. pp. 1–16.

Mayer, Richard E. 2005c. “Cognitive Theory of Multimedia Learning.” In: Mayer. (ed.). 2005a. pp. 31–48.

Mayer, Richard E. 2005d. “Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles.” In: Mayer. (ed.). 2005a. pp. 183–200.

Mayer, Richard E. 2005e. “Principles of Multimedia Learning Based on Social Cues: Personalization, Voice, and Image Principles.” In: Mayer. (ed.). 2005a. pp. 201–212.

Mayer, Richard E. / Johnson, W. Lewis / Shaw, Erin / Sandhu, Sahiba. 2006. “Constructing Computer-Based Tutors that Are Socially Sensitive: Politeness in Educational Software.” In: International Journal of Human-Computer Studies. 64(1). January 2006. pp. 36–42.

Mayer, Richard E. / Moreno, Roxana. 2003. “Nine Ways to Reduce Cognitive Load in Multimedia Learning.” In: Educational Psychologist. 38(1). March 2003. pp. 43–52.

Mayer, Richard E. / Sobko, Kristina / Mautone, Patricia D. 2003. “Social Cues in Multimedia Learning: Role of Speaker’s Voice.” In: Journal of Educational Psychology. 95(2). June 2003. pp. 419–425.

Mayo, Michael J. 2001. “Bayesian Student Modelling and Decision-Theoretic Selection of Tutorial Actions in Intelligent Tutoring Systems.” Ph.D. Thesis. Department of Computer Science and Software Engineering, University of Canterbury. Christchurch, New Zealand. January 2001. Available Online. URL: http://www.cosc.canterbury.ac.nz/research/reports/PhdTheses/2001/phd_0102.pdf. Last Access: 9 September 2009.

Mayo, Michael J. / Mitrovic, Antonija. 2001. “Optimising ITS Behaviour with Bayesian Networks and Decision Theory.” In: International Journal of Artificial Intelligence in Education. 12(2). 2001. pp. 124–153.

McBreen, Helen M. / Jack, Mervyn A. 2000. “Empirical Evaluation of Animated Agents in a Multi-Modal E-Retail Application.” In: Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop. North Falmouth, Massachusetts, USA. 3–5 November 2000.

McBreen, Helen M. / Jack, Mervyn A. 2001. “Evaluating Humanoid Synthetic Agents in E-Retail Applications.” In: IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans. 31(5). September 2001. pp. 394–405.

McCalla, Gordon I. 1992. “The Central Importance of Student Modelling to Intelligent Tutoring.” In: Costa. (ed.). 1992. pp. 107–131.

McCalla, Gordon I. / Vassileva, Julita / Greer, Jim E. / Bull, Susan. 2000. “Active Learner Modelling.” In: Proceedings of the 5th International Conference on Intelligent Tutoring Systems (ITS 2000). Montreal, Quebec, Canada. 19–23 June 2000.

McCarthy, John. 1958. “Programs with Common Sense.” In: Proceedings of the Symposium on Mechanization of Thought Processes. National Physical Laboratory. Teddington, United Kingdom. November 1958.

McCloud, Scott. 1993. Understanding Comics: The Invisible Art. Northampton, Massachusetts: Kitchen Sink Press.

McClurg-Genevese, Joshua D. 2005a. “The Principles of Design.” In: Digital Web Magazine. 13 June 2005. Available Online. URL: http://www.digital-web.com/articles/principles_of_design/. Last Access: 9 September 2009.

McClurg-Genevese, Joshua D. 2005b. “The Elements of Design.” In: Digital Web Magazine. 15 August 2005. Available Online. URL: http://www.digital-web.com/articles/elements_of_design/. Last Access: 9 September 2009.

McCrae, Robert R. / John, Oliver P. 1992. “An Introduction to the Five Factor Model and Its Applications.” In: Journal of Personality. 60(2). June 1992. pp. 175–215.

McDonald, David D. 1992. “Natural Language Generation.” In: Shapiro (ed.). 1992. Volume 2. pp. 983–997.

McFarlane, Daniel C. 1999. “Coordinating the Interruption of People in Human-Computer Interaction.” In: Proceedings of the 7th IFIP Conference on Human-Computer Interaction (INTERACT ’99). Edinburgh, United Kingdom. 30 August–3 September 1999.

McIntosh, Daniel N. 1996. “Facial Feedback Hypotheses: Evidence, Implications, and Directions.” In: Motivation and Emotion. 20(2). June 1996. pp. 121–147.

McKendree, Jean. 1990. “Effective Feedback Content for Tutoring Complex Skills.” In: Human-Computer Interaction. 5(4). December 1990. pp. 381–413.

McKeown, Kathleen R. 1985. “Discourse Strategies for Generating Natural-Language Text.” In: Artificial Intelligence. 27(1). September 1985. pp. 1–41.

McKeown, Kathleen R. / Moore, Johanna D. 1997. “Spoken Language Generation.” In: Cole et al. (eds.). 1997. pp. 182–187.

McKevitt, Paul / Nualláin, Séan Ó. / Mulvihill, Conn. (eds.). 2002. Language, Vision and Music: Selected Papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing (CSNLP-8). Galway, Ireland. 9–11 August 1999. Amsterdam: John Benjamins.

McLaughlin, Margaret L. (ed.). 1987. Communication Yearbook. Volume 10. Newbury Park, California: SAGE Publications.

McLoughlin, Catherine / Oliver, Ron. 1998. “Maximising the Language and Learning Link in Computer Learning Environments.” In: British Journal of Educational Technology. 29(2). April 1998. pp. 125–136.

McMahon, Mark. 1997. “Social Constructivism and the World Wide Web: A Paradigm for Learning.” In: Proceedings of the 14th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE ’97). Perth, Australia. 7–10 December 1997.

McMillan, Sally J. 1998. “Who Pays for Content?: Funding in Interactive Media.” In: Journal of Computer-Mediated Communication. 4(1). September 1998. Available Online. URL: http://jcmc.indiana .edu/vol4/issue1/mcmillan.html. Last Access: 9 September 2009.

McMillan, Sally J. 2000. “Interactivity Is in the Eye of the Beholder: Function, Perception, Involvement, and Attitude Toward the Web Site.” In: Proceedings of the Annual Conference of the American Academy of Advertising. Newport, Rhode Island, USA. 14–17 April 2000.

McMillan, Sally J. 2002. “Exploring Models of Interactivity from Multiple Research Traditions: Users, Documents, and Systems.” In: Lievrouw & Livingston. (eds.). 2002. pp. 162–182.

McMillan, Sally J. / Hwang, Jang-Sun. 2002. “Measures of Perceived Interactivity: An Exploration of the Role of Direction of Communication, User Control, and Time in Shaping Perceptions of Interactivity.” In: Journal of Advertising. 31(3). Fall 2002. pp. 29–42.

McMillan, Sally J. / Hwang, Jang-Sun / Lee, Guiohk. 2003. “Effects of Structural and Perceptual Factors on Attitudes Toward the Website.” In: Journal of Advertising Research. 43(4). December 2003. pp. 400–409.

McNamara, Danielle S. / Levinstein, Irwin B. / Boonthum, Chutima. 2004. “iSTART: Interactive Strategy Training for Active Reading and Thinking.” In: Behavior Research Methods, Instruments, and Computers. 36(2). May 2004. pp. 222–233.

McQuiggan, Scott W. / Robison, Jennifer L. / Phillips, Robert / Lester, James C. 2008. “Modeling Parallel and Reactive Empathy in Virtual Agents: An Inductive Approach.” In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2008). Estoril, Portugal. 12–16 May 2008.

McTear, Michael F. 1993. “User Modelling for Adaptive Computer Systems: A Survey of Recent Developments.” In: Artificial Intelligence Review. 7(3–4). August 1993. pp. 157–184.

McTear, Michael F. 2002. “Spoken Dialogue Technology: Enabling the Conversational Interface.” In: ACM Computing Surveys. 34(1). March 2002. pp. 90–169.

McTear, Michael F. 2004. Spoken Dialogue Technology. London: Springer Verlag.

McTear, Michael F. / O’Neill, Ian / Hanna, Philip / Liu, Xingkun. 2005. “Handling Errors and Determining Confirmation Strategies: An Object-Based Approach.” In: Speech Communication. 45(3). March 2005. pp. 249–269.

Mehrabian, Albert. 1967. “Orientation Behaviors and Nonverbal Attitude Communication.” In: Journal of Communication. 17(4). December 1967. pp. 324–332.

Mehrabian, Albert. 1971. Silent Messages. Belmont, California: Wadsworth Publishing.

Meier, Rolf. 2006. Praxis E-Learning: Grundlagen, Didaktik, Rahmenanalyse, Medienauswahl, Qualifizierungskonzept, Betreuungskonzept, Einführungsstrategie, Erfolgssicherung. Offenbach: GABAL Verlag.

Mellish, Christopher S. / Scott, Donia / Cahill, Lynn / Evans, Roger / Paiva, Daniel S. / Reape, Mike. 2006. “A Reference Architecture for Natural Language Generation Systems.” Technical Report 2006/05. Department of Computing, Faculty of Mathematics and Computing, The Open University. Milton Keynes, United Kingdom. 8 June 2006. Available Online. URL: http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.102.8275&rep=rep1&type=pdf. Last Access: 9 September 2009.

Menzel, Peter / D’Aluisio, Faith. 2000. Robo Sapiens: Evolution of a New Species. Cambridge, Massachusetts: The MIT Press.

Merrill, M. David. 1992. “Constructivism and Instructional Design.” In: Duffy & Jonassen. (eds.). 1992. pp. 99–114.

Metzger, Christiane / Schulmeister, Rolf. 2004. “Interaktivität im Virtuellen Lernen am Beispiel von Lernprogrammen zur Deutschen Gebärdensprache.” In: Mayer & Treichel. (eds.). 2004. pp. 265–297.

Michel, Melchi M. / Jacobs, Robert A. 2007. “Parameter Learning But Not Structure Learning: A Bayesian Network Model of Constraints on Early Perceptual Learning.” In: Journal of Vision. 7(1). January 2007. pp. 1–18.

Middleton, Stuart E. 2001. “Interface Agents: A Review of the Field.” Technical Report ECSTR-IAM01-001. University of Southampton. Southampton, United Kingdom. August 2001. Available Online. URL: http://eprints.ecs.soton.ac.uk/6280/1/agent-survey.pdf. Last Access: 9 September 2009.

Miles, Ian 1992. “When Mediation Is the Message: How Suppliers Envisage New Markets.” In: Lea. (ed.). 1992. pp. 145–167.

Miles, Matthew B. / Huberman, A. Michael. 1994. Qualitative Data Analysis: An Expanded Sourcebook (2nd ed.). Thousand Oaks, California: SAGE Publications.

Milgram, Stanley. 1974. Obedience to Authority. London: Tavistock.

Miller, George A. 1956. “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.” In: Psychological Review. 63(2). March 1956. pp. 81–97.

Mills, Sara. 2003. Gender and Politeness. Cambridge: Cambridge University Press.

Milton, Nicholas R. 2007. Knowledge Acquisition in Practice: A Step-by-Step Guide. London: Springer Verlag.

Minato, Takashi / Shimada, Michihiro / Ishiguro, Hiroshi / Itakura, Shoji. 2004. “Development of an Android Robot for Studying Human-Robot Interaction.” In: Proceedings of the 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2004). Ottawa, Ontario, Canada. 17–20 May 2004.

Minsky, Marvin. 1975. “A Framework for Representing Knowledge.” In: Winston. (ed.). 1975. pp. 211–277.

Minsky, Marvin. 1982. “Why People Think Computers Can’t.” In: AI Magazine 3(4). Fall 1982. pp. 3–15.

Minsky, Marvin. 1985. The Society of Mind. New York: Simon and Schuster.

Mintzes, Joel J. / Wandersee, James H. / Novak, Joseph D. (eds.). 1998. Teaching Science for Understanding: A Human Constructivist View. San Diego: Academic Press.

Mishra, Punyashloke. 2006. “Affective Feedback from Computers and Its Effect on Perceived Ability and Affect: A Test of the Computers as Social Actor Hypothesis.” In: Journal of Educational Multimedia and Hypermedia. 15(1). January 2006. pp. 107–131.

Mishra, Punyashloke / Nicholson, Michael D. / Wojcikiewicz, Steven. 2001. “Does My Word Processor Have a Personality?: Or Topffer’s Law and Educational Technology.” In: Journal of Adolescent and Adult Literacy. 44 (7). April 2001. pp. 634–641.

Mishra, Punyashloke / Wojcikiewicz, Steven. 2002. “Taking Things at Face Value: The Psychology of Media Response.” Review of Reeves & Nass (1998). In: Journal of Educational Computing Research. 26(2). 2002. pp. 219–226.

Mishra, Punyashloke / Zhao, Yong / The Social Interfaces and Learning Technologies (SILT) Group. 2000. “Introduction: Social Responses to Interactive Media.” In: “Psychological Responses to Interactive Technology: Exploring the Educational Implications of the Media Equation.” Symposium at the Annual Conference of the American Educational Research Association (AERA 2000). New Orleans, Louisiana, USA. 24–28 April 2000.

Mitchell, Tom M. 1997. Machine Learning. Columbus, Ohio: McGraw-Hill.

Mitkov, Ruslan. (ed.). 2003. The Oxford Handbook of Computational Linguistics. Oxford: Oxford University Press.

Mitrovic, Antonija / Koedinger, Kenneth R. / Martin, Brent. 2003. “A Comparative Analysis of Cognitive Tutoring and Constraint-Based Modeling.” In: Proceedings of the 9th International Conference on User Modeling (UM 2003). Johnstown, Pennsylvania, USA. 22–26 June 2003.

Mitrovic, Antonija / Mayo, Michael J. / Suraweera, Pramuditha / Martin, Brent. 2001. “Constraint-Based Tutors: A Success Story.” In: Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2001). Budapest, Hungary. 4–7 June 2001.

Mitrovic, Antonija / Ohlsson, Stellan. 1999. “Evaluation of a Constraint-Based Tutor for a Database Language.” In: International Journal of Artificial Intelligence in Education. 10(3–4). 1999. pp. 238–256.

Mitrovic, Antonija / Ohlsson, Stellan / Martin, Brent. 2006. “Problem Solving Support in Constraint-Based Tutors.” In: Technology, Instruction, Cognition and Learning (TICL). 3(1–2). 2006. pp. 43–50.

Moffat, Dave. 1997. “Personality Parameters and Programs.” In: Trappl & Petta. (eds.). 1997. pp. 120–165.

Moon, Youngme. 1998. “Intimate Self-Disclosure Exchanges: Using Computers to Build Reciprocal Relationships with Consumers.” Working Paper 99-059. Harvard Business School. Cambridge, Massachusetts, USA. November 1998.

Moon, Youngme / Nass, Clifford I. 1996. “How ‘Real’ Are Computer Personalities? Psychological Responses to Personality Types in Human-Computer Interaction.” In: Communication Research. 23(6). December 1996. pp. 651–674.

Mooney, Raymond J. 2003. “Machine Learning.” In: Mitkov. (ed.). 2003. pp. 376–394.

Moore, Michael G. 1972. “Learner Autonomy: The Second Dimension of Independent Learning.” In: Convergence. 5(2). Fall 1972. pp. 76–88.

Moore, Michael G. 1989. “Three Types of Interaction.” In: The American Journal of Distance Education. 3(2). 1989. pp. 1–6.

Moore, Michael G. 1991. “Distance Education Theory.” In: The American Journal of Distance Education. 5(3). 1991. pp. 1–6.

Moore, Michael G. 1993. “Theory of Transactional Distance.” In: Keegan. (ed.). 1993. pp. 22–38.

Moore, Michael G. / Kearsley, Greg. 1996. Distance Education: A Systems View. Belmont, California: Wadsworth Publishing.

Moore, Michael G. / Anderson, William G. (eds.). 2003. Handbook of Distance Education. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Morel, Benoît. 2004. “Recruiting a Virtual Employee: Adaptive and Personalized Agents in Corporate Communication.” In: Payr & Trappl. (eds.). 2004. pp. 177–196.

Moreno, Kristen N. / Person, Natalie K. / Adcock, Amy B. / Van Eck, Richard N. / Jackson, G. Tanner / Marineau, Johanna C. 2002. “Etiquette and Efficacy in Animated Pedagogical Agents: The Role of Stereotypes.” In: Proceedings of the AAAI Fall Symposium on Personalized Agents. North Falmouth, Massachusetts, USA. 15–17 November 2002.

Moreno, Roxana. 2005. “Instructional Technology: Promise and Pitfalls.” In: PytlikZillig et al. (eds.). 2005. pp. 1–19.

Moreno, Roxana / Flowerday, Terri. 2006. “Students’ Choice of Animated Pedagogical Agents in Science Learning: A Test of the Similarity-Attraction Hypothesis on Gender and Ethnicity.” In: Contemporary Educational Psychology. 31(2). April 2006. pp. 186–207.

Moreno, Roxana / Mayer, Richard E. 1999. “Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity.” In: Journal of Educational Psychology. 91(2). June 1999. pp. 358–368.

Moreno, Roxana / Mayer, Richard E. 2000. “Pedagogical Agents in Constructivist Multimedia Environments: The Role of Image and Language in the Instructional Communication.” Paper Presented at the Annual Conference of the American Educational Research Association (AERA 2000). New Orleans, Louisiana, USA. 24–28 April 2000.

Moreno, Roxana / Mayer, Richard E. 2007. “Interactive Multimodal Learning Environments.” In: Educational Psychology Review. 19(3). September 2007. pp. 309–326.

Moreno, Roxana / Mayer, Richard E. / Lester, James C. 2000. “Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of Their Interaction.” In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2000). Montreal, Quebec, Canada. 26 June–1 July 2000.

Moreno, Roxana / Mayer, Richard E. / Spires, Hiller A. / Lester, James C. 2001. “The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact with Animated Pedagogical Agents?.” In: Cognition & Instruction. 19(2). 2001. pp. 177–213.

Morgan, David L. 1997. Focus Groups as Qualitative Research (2nd ed.). London: SAGE Publications.

Morgan, Jeffrey J. 1995. “Anthropomorphism on Trial.” Submission for Interacting with Computers. Available Online. URL: http://citeseer.ist.psu.edu/morgan95submission.html. Last Access: 9 September 2009.

Mori, Masahiro. 1970. “Bukimi No Tani (The Uncanny Valley).” In: Energy. 7(4). 1970. pp. 33–35. English Translation by Karl F. MacDorman and Takashi Minato. Available Online. URL: http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html. Last Access: 9 September 2009.

Mori, Masahiro. 1982. The Buddha in the Robot. Boston, Massachusetts: Charles E. Tuttle.

Morkes, John / Kernal, Hadyn K. / Nass, Clifford I. 1999. “Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory.” In: Human-Computer Interaction. 14(4). December 1999. pp. 395–435.

Morra-Imas, Linda / Rist, Ray C. 2002. “Building Skills to Evaluate Development Interventions, Module 8: Data Analysis and Interpretation.” International Program for Development Evaluation Training (IPDET). The World Bank Group. Washington, District of Columbia, USA. / Carleton University. Ottawa, Ontario, Canada. / IOB/Ministry of Foreign Affairs. The Hague, The Netherlands. Available Online. URL: http://www.insp.mx/bidimasp/bibliografia/1/6/IPDET%20modulo%208.pdf. Last Access: 10 September 2009.

Mostow, Jack / Aist, Gregory. 2001. “Evaluating Tutors that Listen: An Overview of Project LISTEN.” In: Forbus & Feltovich. (eds.). 2001. pp. 169–234.

Mostow, Jack / Beck, Joseph. 2007. "When the Rubber Meets the Road: Lessons from the In-School Adventures of an Automated Reading Tutor that Listens." In: Schneider & McDonald. (eds.). 2007. pp. 183–200.

Motta, Enrico. 2006. "Knowledge Publishing and Access on the Semantic Web: A Sociotechnological Analysis." In: IEEE Intelligent Systems. May–June 2006. pp. 88–90.

Moundridou, Maria / Virvou, Maria. 2002. "Evaluating the Persona Effect of an Interface Agent in an Intelligent Tutoring System." In: Journal of Computer Assisted Learning. 18(3). September 2002. pp. 253–261.

Muirhead, Brent / Juwah, Charles. 2004. "Interactivity in Computer-Mediated College and University Education: A Recent Review of the Literature." In: Journal of Educational Technology & Society. 7(1). January 2004. pp. 12–20.

Mulder, Matthijs P. / Nijholt, Anton. 2002. "Humour Research: State of the Art." Technical Report CTIT-02-34. Center for Telematics and Information Technology (CTIT), University of Twente. Enschede, The Netherlands. September 2002. Available Online. URL: http://doc.utwente.nl/38233/1/0000009e.pdf. Last Access: 10 September 2009.

Mullennix, John W. / Stern, Steven E. / Wilson, Stephen J. / Dyson, Corrie-Lynn. 2003. "Social Perception of Male and Female Computer Synthesized Speech." In: Computers in Human Behavior. 19(4). July 2003. pp. 407–424.

Muller, Michael J. / Kuhn, Sarah. (eds.). 1993. Communications of the ACM: Special Issue on Participatory Design. 36(4). June 1993.

Muller, Michael J. / Haslwanter, Jean D. H. / Dayton, Thomas. 1997. "Participatory Practices in the Software Lifecycle." In: Helander et al. (eds.). 1997. pp. 255–297.

Murray, Iain R. / Arnott, John L. 1993. "Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion." In: Journal of the Acoustic Society of America. 93(2). February 1993. pp. 1097–1108.

Murray, Janet H. 1998. Hamlet on the Holodeck: The Future of Narrative in Cyberspace. Cambridge, Massachusetts: The MIT Press.

Murray, R. Charles / VanLehn, Kurt / Mostow, Jack. 2001. "A Decision-Theoretic Architecture for Selecting Tutorial Discourse Actions." In: Proceedings of the Workshop on Tutorial Dialog Systems. 10th International Conference on Artificial Intelligence in Education (AI-ED 2001). San Antonio, Texas, USA. 19–23 May 2001.

Murray, R. Charles / VanLehn, Kurt / Mostow, Jack. 2004. "Looking Ahead to Select Tutorial Actions: A Decision-Theoretic Approach." In: International Journal of Artificial Intelligence in Education. 14(3–4). 2004. pp. 235–278.

Myllymäki, Petri. 2005. "Advantages of Bayesian Networks in Data Mining and Knowledge Discovery." Bayes Information Technology Ltd. Helsinki, Finland. Available Online. URL: http://www.bayesit.com/docs/advantages.html. Last Access: 10 September 2009.

Nadel, Lynn. (ed.). 2003. Encyclopedia of Cognitive Science. 4 Volumes. London: Nature Publishing Group.

Naismith, Laura / Lonsdale, Peter / Vavoula, Giasemi / Sharples, Mike. 2004. "Literature Review in Mobile Technologies and Learning." Report 11. Futurelab Series. University of Birmingham. Birmingham, United Kingdom. December 2004. Available Online. URL: http://www.futurelab.org.uk/resources/documents/lit_reviews/Mobile_Review.pdf. Last Access: 10 September 2009.

Narayek, Alexander. 2004. “Artificial Intelligence in Computer Games: State of the Art and Future Directions.” In: ACM Queue. 1(10). February 2004. pp. 58–65.

Nass, Clifford I. / Brave, Scott. 2005. Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. Cambridge, Massachusetts: The MIT Press.

Nass, Clifford I. / Fogg, Brian J. / Moon, Youngme. 1996. “Can Computers be Teammates?.” In: International Journal of Human-Computer Studies. 45(6). December 1996. pp. 669–678.

Nass, Clifford I. / Gong, Li. 1999. “Maximized Modality or Constrained Consistency?.” In: Proceedings of the International Conference on Audio-Visual Speech Processing (AVSP ’99). Santa Cruz, California, USA. 7–9 August 1999.

Nass, Clifford I. / Isbister, Katherine / Lee, Eun-Ju. 2000a. “Truth Is Beauty: Researching Embodied Conversational Agents.” In: Cassell et al. (eds.). 2000a. pp. 374–402.

Nass, Clifford I. / Simard, Caroline / Takhteyev, Yuri. 2000b. “Should Recorded and Synthesized Speech Be Mixed?.” Department of Communication, Stanford University. Stanford, California, USA. Available Online. URL: http://www.stanford.edu/~nass/comm369/pdf/MixingTTSandRecordedSpeech .pdf. Last Access: 10 September 2009.

Nass, Clifford I. / Kim, Eun-Young / Lee, Eun-Ju. 1998. “When My Face Is the Interface: An Experimental Comparison of Interacting with One’s Own Face or Someone Else’s Face.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI ’98). Los Angeles, California, USA. 18–23 April 1998.

Nass, Clifford I. / Lee, Kwan Min. 2001. “Does Computer-Synthesized Speech Manifest Personality?: Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction.” In: Journal of Experimental Psychology: Applied. 7(3). September 2001. pp. 171–181.

Nass, Clifford I. / Moon, Youngme. 2000. “Machines and Mindlessness: Social Responses to Computers.” In: Journal of Social Issues. 56(1). Spring 2000. pp. 81–103.

Nass, Clifford I. / Moon, Youngme / Carney, Paul. 1999. “Are People Polite to Computers?: Responses to Computer-Based Interviewing Systems.” In: Journal of Applied Social Psychology. 29(5). May 1999. pp. 1093–1109.

Nass, Clifford I. / Moon, Youngme / Fogg, Brian J. / Reeves, Byron / Dryer, D. Christopher. 1995. “Can Computer Personalities be Human Personalities?.” In: International Journal of Human-Computer Studies. 43(2). August 1995. pp. 223–239.

Nass, Clifford I. / Moon, Youngme / Green, Nancy. 1997a. “Are Machines Gender Neutral?: Gender-Stereotypic Responses to Computers with Voices.” In: Journal of Applied Social Psychology. 27(10). 16–31 May 1997. pp. 864–876.

Nass, Clifford I. / Moon, Youngme / Morkes, John / Kim, Eun-Young / Fogg, Brian J. 1997b. “Computers Are Social Actors: A Review of Current Research.” In: Friedman. (ed.). 1997. pp. 137–162.

Nass, Clifford I. / Robles, Erica / Wang, Qianying. 2004. “‘User as Assessor’ Approach to Embodied Conversational Agents.” In: Ruttkay & Pelachaud. (eds.). 2004. pp. 161–188.

Nass, Clifford I. / Steuer, Jonathan. 1993. “Voices, Boxes, and Sources of Messages: Computers and Social Actors.” In: Human Communication Research. 19(4). June 1993. pp. 504–527.

Nass, Clifford I. / Steuer, Jonathan / Tauber, Ellen R. 1994. “Computers Are Social Actors.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI ’94). Boston, Massachusetts, USA. 24–28 April 1994.

Nass, Clifford I. / Sundar, S. Shyam. 1994. "Is Human-Computer Interaction Social or Parasocial?." Department of Communication, Stanford University. Stanford, California, USA. Submitted to Human Communication Research. 17 August 1994. Available Online. URL: http://www.stanford.edu/group/commdept/oldstuff/srct_pages/Social-Parasocial.html. Last Access: 10 September 2009.

Nass, Clifford I. / Takayama, Leila / Brave, Scott. 2006. "Socializing Consistency: From Technical Homogeneity to Human Epitome." In: Zhang & Galletta. (eds.). 2006. pp. 373–392.

Neal, Lisa / Miller, Diane. 2005. "Distance Education." In: Proctor & Vu. (eds.). 2005. pp. 454–470.

Nebel, Bernhard. 2001. "Frame-Based Systems." In: Wilson & Keil. (eds.). 2001. pp. 324–326.

Negroponte, Nicholas. 1995. Being Digital. London: Hodder and Stoughton.

Neisser, Ulrich. 1967. Cognitive Psychology. New York: Appleton-Century-Crofts.

Nelson, Theodor H. 1965. "Complex Information Processing: A File Structure for the Complex, the Changing and the Indeterminate." In: Proceedings of the 20th Annual ACM National Conference. Cleveland, Ohio, USA. 24–26 August 1965.

Neo, Mai / Neo, Ken T. K. 2001. "Innovative Teaching: Using Multimedia in a Problem-Based Learning Environment." In: Journal of Educational Technology & Society. 4(4). October 2001. pp. 19–31.

Nerbonne, John. 2003. "Natural Language Processing in Computer-Assisted Language Learning." In: Mitkov. (ed.). 2003. pp. 670–698.

Neumann, Bernd. 2000. "Praktische Informatik III (P3), Teil 4: Agententechnologie." Lecture Notes. Department of Computer Science, University of Hamburg. Hamburg, Germany. Winter Term 2000. Available Online. URL: http://kogs-www.informatik.uni-hamburg.de/~neumann/P3-WS-2000/P3-Teil4.pdf. Last Access: 17 September 2009.

Neuman, W. Russell. 1991. The Future of the Mass Audience. New York: Cambridge University Press.

Neumann, Eric K. / Miller, Eric / Wilbanks, John. 2004. "What the Semantic Web Could Do for the Life Sciences." In: Drug Discovery Today: BIOSILICO. 2(6). November 2004. pp. 228–236.

Newell, Allen / Simon, Herbert A. 1976. "Computer Science as Empirical Inquiry: Symbols and Search." In: Communications of the ACM. 19(3). March 1976. pp. 113–126.

Newhagen, John E. 2004. "Interactivity, Dynamic Symbol Processing, and the Emergence of Content in Human Communication." In: The Information Society. 20(5). November–December 2004. pp. 395–400.

Newhagen, John E. / Cordes, John W. / Levy, Mark R. 1995. "Nightly@nbc.com: Audience Scope and the Perception of Interactivity in Viewer Mail on the Internet." In: Journal of Communication. 45(3). September 1995. pp. 164–175.

Nguyen, Hien / Masthoff, Judith / Edwards, Pete. 2007. "Persuasive Effects of Embodied Conversational Agent Teams." In: Proceedings of the 12th International Conference on Human-Computer Interaction (HCI International 2007). Beijing, China. 22–27 July 2007.

Nielsen, Jakob. 1994. "Heuristic Evaluation." In: Nielsen & Mack. (eds.). 1994. pp. 25–62.

Nielsen, Jakob / Mack, Robert L. (eds.). 1994. Usability Inspection Methods. New York: John Wiley & Sons.

Nijholt, Anton. 2003. "Humor and Embodied Conversational Agents." Technical Report TR-CTIT-03-03. Center for Telematics and Information Technology (CTIT), University of Twente. Enschede, The Netherlands. March 2003. Available Online. URL: http://www.ub.utwente.nl/webdocs/ctit/1/000000b3.pdf. Last Access: 10 September 2009.

Nijholt, Anton. 2006. "Embodied Conversational Agents: 'A Little Humor Too'." In: IEEE Intelligent Systems. 21(2). March–April 2006. pp. 62–64.

Nilsson, Nils J. 1996. "Introduction to Machine Learning." Incomplete Draft of Textbook. Artificial Intelligence Laboratory, Department of Computer Science, Stanford University. Stanford, California, USA. 4 December 1996. Available Online. URL: http://ai.stanford.edu/people/nilsson/MLDraftBook/MLBOOK.pdf. Last Access: 10 September 2009.

Nimmo, Dan D. (ed.). 1979. Communication Yearbook. Volume 3. New Brunswick, New Jersey: Transaction Books.

Noot, Han / Ruttkay, Zsófia. 2005. "Variations in Gesturing and Speech by GESTYLE." In: International Journal of Human-Computer Studies. 62(2). February 2005. pp. 211–229.

Norman, Donald A. 1988, 2002a. The Design of Everyday Things. New York: Basic Books.

Norman, Donald A. 2002b. "Emotion & Design: Attractive Things Work Better." In: Interactions. 9(4). July 2002. pp. 36–42.

Norman, Donald A. 1997. "How Might People Interact with Agents." In: Bradshaw. (ed.). 1997. pp. 49–56.

Norman, Donald A. 1998. The Invisible Computer. Cambridge, Massachusetts: The MIT Press.

Norman, Donald A. 2004. Emotional Design. New York: Basic Books.

Norman, Donald A. / Ortony, Andrew / Russell, Daniel M. 2003. "Affect and Machine Design: Lessons for the Development of Autonomous Machines." In: IBM Systems Journal. 42(1). 2003. pp. 38–44.

Northrup, Pam. 2001. "A Framework for Designing Interactivity into Web-Based Instruction." In: Educational Technology. 41(2). March–April 2001. pp. 31–39.

Northrup, Pam / Lee, Russell / Burgess, Vance. 2002. "Learner Perceptions of Online Interaction." In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2002). Denver, Colorado, USA. 24–29 June 2002.

Novak, Joseph D. / Cañas, Alberto J. 2006. "The Theory Underlying Concept Maps and How to Construct and Use Them." Technical Report IHMC CmapTools 2006-01. Florida Institute for Human and Machine Cognition (IHMC). Pensacola, Florida, USA. Available Online. URL: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMapsHQ.pdf. Last Access: 10 September 2009.

Novak, Thomas P. / Hoffman, Donna L. / Yung, Yiu-Fai. 2000. "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach." In: Marketing Science. 19(1). Special Issue on Marketing Science and the Internet. Winter 2000. pp. 22–42.

Nowak, Kristine L. 2004. "The Influence of Anthropomorphism and Agency on Social Judgment in Virtual Environments." In: Journal of Computer-Mediated Communication. 9(2). January 2004. Available Online. URL: http://jcmc.indiana.edu/vol9/issue2/nowak.html. Last Access: 10 September 2009.

Nowak, Kristine L. / Biocca, Frank. 2003. "The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments." In: Presence: Teleoperators and Virtual Environments. 12(5). October 2003. pp. 481–494.

Nwana, Hyacinth S. 1991. "User Modelling and User-Adapted Interaction in an Intelligent Tutoring System." In: User Modeling and User-Adapted Interaction. 1(1). March 1991. pp. 1–32.

Nwana, Hyacinth S. 1996. "Software Agents: An Overview." In: The Knowledge Engineering Review. 11(3). September 1996. pp 205–244.

Oblinger, Diana G. / Oblinger, James L. (eds.). 2005. Educating the Net Generation. EDUCAUSE. Available Online. URL: http://net.educause.edu/ir/library/pdf/pub7101.pdf. Last Access: 10 September 2009.

Odell, James / Van Dyke Parunak, Harry / Fleischer, Mitchell. 2003. "The Role of Roles." In: Journal of Object Technology. 2(1). January–February 2003. pp. 39–51.

OED. 1998. "The New Oxford Dictionary of English." Edited by Judy Pearsall. Oxford: Oxford University Press.

Ohlsson, Stellan. 1992. "Constraint-Based Student Modelling." In: International Journal of Artificial Intelligence in Education. 3(4). 1992. pp. 429–447.

Ohlsson, Stellan. 1994. "Constraint-Based Student Modelling." In: Greer & McCalla. (eds.). 1994. pp. 167–189.

Ohlsson, Stellan / Mitrovic, Antonija. 2006. "Constraint-Based Knowledge Representation for Individualized Instruction." In: Computer Science and Information Systems. 3(1). June 2006. pp. 1–22.

Oikonomopoulos, Antonios / Patras, Ioannis / Pantic, Maja. 2006. "Spatiotemporal Salient Points for Visual Recognition of Human Actions." In: IEEE Transactions on Systems, Man, and Cybernetics: Part B: Cybernetics. 36(3). June 2006. pp. 710–719.

O'Malley, Claire / Vavoula, Giasemi N. / Glew, J. Phil / Taylor, Josie / Sharples, Mike / Lefrere, Paul. 2003. "Guidelines for Learning/Teaching/Tutoring in a Mobile Environment." Report MOBIlearn/UoN, UoB, OU/D4.1/1.0. 10 June 2003. Available Online. URL: http://www.mobilearn.org/download/results/guidelines.pdf. Last Access: 10 September 2009.

Opperman, Reinhard. (ed.). 1994. Adaptive User Support: Ergonomic Design of Manually and Automatically Adaptable Software. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

O'Reilly, Tenaha / Sinclair, Grant P. / McNamara, Danielle S. 2004. "iSTART: A Web-Based Reading Strategy Intervention that Improves Students' Science Comprehension." In: Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2004). Lisbon, Portugal. 15–17 December 2004.

O'Reilly, Tim. 2005. "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software." O'Reilly Media. 30 September 2005. Available Online. URL: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html. Last Access: 10 September 2009.

Ormrod, Jeanne E. 2003. Educational Psychology: Developing Learners (4th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Ortony, Andrew. 2003. "On Making Believable Emotional Agents Believable." In: Trappl et al. (eds.). 2003. pp. 189–211.

Ortony, Andrew / Clore, Gerald L. / Collins, Allan M. 1988. The Cognitive Structure of Emotions. Cambridge: Cambridge University Press.

Ortony, Andrew / Norman, Donald A. / Revelle, William. 2005. "Affect and Proto-Affect in Effective Functioning." In: Fellous & Arbib. (eds.). 2005. pp. 173–202.

Ortony, Andrew / Turner, Terence J. 1990. "What's Basic About Basic Emotions?." In: Psychological Review. 97(3). July 1990. pp. 315–331.

Oudeyer, Pierre-Yves. 2003. "The Production and Recognition of Emotions in Speech: Features and Algorithms." In: International Journal of Human-Computer Studies. 59(1–2). July 2003. pp. 157–183.

Oviatt, Sharon L. 1999. “Ten Myths of Multimodal Interaction.” In: Communications of the ACM. 42(11). November 1999. pp. 74–81.

Oviatt, Sharon L. 2000. “Taming Recognition Errors with a Multimodal Interface.” In: Communications of the ACM. 43(9). September 2000. pp. 45–51.

Oviatt, Sharon L. 2003. “Multimodal Interfaces.” In: Jacko & Sears. (eds.). 2003. pp. 286–304.

Oviatt, Sharon L. / Adams, Bridget. 2000. “Designing and Evaluating Conversational Interfaces with Animated Characters.” In: Cassell et al. (eds.). 2000a. pp. 319–345.

Oviatt, Sharon L. / Cohen, Philip. 2000. “Multimodal Interfaces that Process What Comes Naturally.” In: Communications of the ACM. 43(3). March 2000. pp. 45–53.

Oviatt, Sharon L. / Cohen, Philip / Wu, Lizhong / Vergo, John / Duncan, Lisbeth / Suhm, Bernhard / Bers, Josh / Holzman, Thomas / Winograd, Terry / Landay, James / Larson, Jim / Ferro, David. 2000. “Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions.” In: Human-Computer Interaction. 15(4). December 2000. pp. 263–322.

Oviatt, Sharon L. / De Angeli, Antonella / Kuhn, Karen. 1997. “Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI ’97). Atlanta, Georgia, USA. 22–27 March 1997.

Paas, Fred / Renkl, Alexander / Sweller, John. 2003. “Cognitive Load Theory and Instructional Design: Recent Developments.” In: Educational Psychologist. 38(1). March 2003. pp. 1–4.

Padayachee, Indira. 2002. “Intelligent Tutoring Systems: Architecture and Characteristics.” In: Proceedings of the 32nd Annual Southern African Computer Lecturer’s Association Conference (SACLA 2002). Fish River Sun, Eastern Cape, South Africa. 27–29 June 2002.

Pagani, Margherita. (ed.). 2005. Encyclopedia of Multimedia Technology and Networking. 2 Volumes. Hersky, Pennsylvania: Idea Group Publishing.

Paiva, Ana / Dias, João / Sobral, Daniel / Aylett, Ruth S. / Sobreperez, Polly / Woods, Sarah / Zoll, Carsten / Hall, Lynne. 2004. “Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents.” In: Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004). New York City, New York, USA. 19–23 July 2004.

Paiva, Ana / Machado, Isabel. 1998. “Vincent, an Autonomous Pedagogical Agent for On-the-Job Training.” In: Proceedings of the 4th International Conference on Intelligent Tutoring Systems (ITS ’98). San Antonio, Texas, USA. 16–19 August 1998.

Paiva, Ana / Machado, Isabel. 2002. “Lifelong Training with Vincent, a Web-Based Pedagogical Agent.” In: International Journal of Continuing Engineering Education and Life Long Learning. 12(1–4). 2002. pp. 254–266.

Paiva, Ana / Machado, Isabel / Martinho, Carlos. 1999. “Enriching Pedagogical Agents with Emotional Behavior: The Case of Vincent.” In: Proceedings of the Workshop on Animated and Personified Pedagogical Agents. 9th International Conference on Artificial Intelligence in Education (AI-ED ’99). Le Mans, France. 19–23 July 1999.

Paiva, Daniel S. 1998. “A Survey of Applied Natural Language Generation Systems.” Technical Report ITRI-98-03. Information Technology Research Institute (ITRI), University of Brighton. Brighton, United Kingdom. July 1998. Available Online. URL: http://www.itri.brighton.ac.uk/projects/rags/surveyAppliedNLG.ps.gz. Last Access: 11 September 2009.

Pan, Shimei. 2002. “Prosody Modeling in Concept-to-Speech Generation.” Ph.D. Thesis. Graduate School of Arts and Sciences, Columbia University. New York City, New York, USA. September 2002. Available Online. URL: http://www1.cs.columbia.edu/nlp/theses/shimei_pan.pdf. Last Access: 11 September 2009.

Panke, Stefanie. 2007. “Unterwegs im Web 2.0: Charakteristiken und Potenziale.” E-teaching.org, Portalbereich Didaktisches Design. 19 January 2007. Available Online. URL: http://www.e-teaching.org/didaktik/theorie/informelleslernen/Web2.pdf. Last Access: 11 September 2009.

Pantic, Maja. 2005. “Face for Interface.” In: Pagani. (ed.). 2005. Volume 1. pp. 308–314.

Pantic, Maja / Bartlett, Marian S. 2007. “Machine Analysis of Facial Expressions.” In: Delac & Grgic. (eds.). 2007. pp. 377–416.

Pantic, Maja / Pentland, Alex P. / Nijholt, Anton / Huang, Thomas S. 2007. “Human Computing and Machine Understanding of Human Behavior: A Survey.” In: Huang et al. (eds.). 2007. pp. 47–71.

Pantic, Maja / Rothkrantz, Leon J. M. 2000. “Automatic Analysis of Facial Expressions: The State of the Art.” In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(12). December 2000. pp. 1424–1445.

Pantic, Maja / Rothkrantz, Leon J. M. 2003. “Toward an Affect-Sensitive Multimodal Human-Computer Interaction.” In: Proceedings of the IEEE. 91(9). September 2003. pp. 1370–1390.

Papert, Seymour A. 1980. Mindstorms: Children, Computers, and Powerful Ideas. New York: Basic Books.

Parise, Salvatore / Kiesler, Sara / Sproull, Lee / Waters, Keith. 1999. “Cooperating with Life-Like Interface Agents.” In: Computers in Human Behavior. 15(2). March 1999. pp. 123–142.

Park, Sung / Catrambone, Richard. 2007. “Understanding the Social Relationship Between Humans and Virtual Humans.” In: Proceedings of the 12th International Conference on Human-Computer Interaction (HCI International 2007). Beijing, China. 22–27 July 2007.

Parke, Frederic I. 1991. “Techniques of Facial Animation.” In: Magnenat-Thalmann & Thalmann. (eds.). 1991. pp. 229–241.

Parsia, Bijan / Sirin, Evren / Grau, Bernardo C. / Ruckhaus, Edna / Hewlett, Daniel. 2005. “Cautiously Approaching SWRL.” Maryland Information and Network Dynamics Lab Semantic Web Agents Project (MINDSWAP), University of Maryland. College Park, Maryland, USA. Available Online. URL: http://www.mindswap.org/papers/CautiousSWRL.pdf. Last Access: 11 September 2009.

Patrick, Andrew S. 2002. “Building Trustworthy Software Agents.” In: IEEE Internet Computing. 6(6). November–December 2002. pp. 46–53.

Paulhus, Delroy L. 1991. “Measurement and Control of Response Bias.” In: Robinson et al. (eds.). 1991. pp. 17–59.

Paulus, Paul B. (ed.). 1983. Basic Group Processes. New York: Springer Verlag.

Pavlik, John V. 1998. New Media Technology: Cultural and Commercial Perspectives (2nd ed.). Boston, Massachusetts: Allyn and Bacon.

Pavlov, Ivan P. 1927. Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. (Translated and Edited by Gleb V. Anrep.). London: Oxford University Press.

Payne, Stephen J. / Squibb, Helen R. 1990. “Algebra Mal-Rules and Cognitive Accounts of Error.” In: Cognitive Science. 14(3). July–September 1990. pp. 445–481.

Payr, Sabine / Trappl, Robert. (eds.). 2004. Agent Culture: Human-Agent Interaction in a Multicultural World. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Pear, Joseph J. 2001. The Science of Learning. Philadelphia, Pennsylvania: Psychology Press.

Pearl, Judea. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco: Morgan Kaufmann Publishers.

Pearl, Judea. 2001. "Bayesian Networks, Causal Inference and Knowledge Discovery." Technical Report R-281. Cognitive Systems Laboratory, University of California at Los Angeles (UCLA). Los Angeles, California, USA. 5 March 2001. Available Online. URL: http://ftp.cs.ucla.edu/pub/stat_ser/R281.pdf. Last Access: 11 September 2009.

Pearl, Judea / Russell, Stuart J. 2003. "Bayesian Networks." In: Arbib. (ed.). 2003. pp. 157–160.

Pearson, Jamie / Hu, Jiang / Branigan, Holly P. / Pickering, Martin J. / Nass, Clifford I. 2006. "Adaptive Language Behavior in HCI: How Expectations and Beliefs About a System Affect Users' Word Choice." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2006). Montreal, Quebec, Canada. 22–27 April 2006.

Peccei, Jean S. 1999. Pragmatics. London: Routledge.

Pelachaud, Catherine. 2003. "Emotion Expressiveness Embedded in Representation Languages for ECAs." Technical Report 2. Preparing Future Multisensorial Interaction Research (PF-STAR) Project. 28 November 2003. Available Online. URL: http://pfstar.itc.it/public/doc/deliverables/pelachaud_tech_rep2.pdf. Last Access: 11 September 2009.

Pelachaud, Catherine / Badler, Norman I. / Steedman, Mark. 1996. "Generating Facial Expressions for Speech." In: Cognitive Science. 20(1). January–March 1996. pp. 1–46.

Pellone, Gennaro. 1995. "Educational Software Design: A Literature Review." In: Australian Journal of Educational Technology. 11(1). Winter 1995. pp. 68–84.

Pentland, Alex P. 1996. "Smart Rooms." In: Scientific American. 274(4). April 1996. pp. 54–62.

Pereira, Fernando. 1997. "Sentence Modeling and Parsing." In: Cole et al. (eds.). 1997. pp. 111–121.

Person, Nathalie K. / Graesser, Arthur C. / Kreuz, Roger J. / Pomeroy, Victoria / The Tutoring Research Group. 2001. "Simulating Human Tutor Dialog Moves in AutoTutor." In: International Journal of Artificial Intelligence in Education. 12(1). 2001. pp. 23–39.

Person, Nathalie K. / Graesser, Arthur C. 2003a. "Pedagogical Agents and Tutors." In: Guthrie. (ed.). 2003. pp. 1169–1172.

Person, Nathalie K. / Graesser, Arthur C. 2003b. "Fourteen Facts About Human Tutoring: Food for Thought for ITS Developers." In: Proceedings of the Workshop "Tutorial Dialogue Systems: With a View Toward the Classroom." 11th International Conference on Artificial Intelligence in Education (AI-ED 2003). Sydney, Australia. 20–24 July 2003.

Pervin, Lawrence A. / John, Oliver P. (eds.). 1999. Handbook of Personality: Theory and Research (2nd ed.). New York: The Guilford Press.

Peterson, Lloyd R. / Peterson, Margaret J. 1959. "Short-Term Retention of Individual Verbal Items." In: Journal of Experimental Psychology. 58(3). September 1959. pp. 193–198.

Petta, Paolo. 2003. "The Role of Emotions in a Tractable Architecture for Situated Cognizers." In: Trappl et al. (eds.). 2003. pp. 251–287.

Pettijohn, Terry F. 1998. Psychology: A ConnecText (4th ed.). Hightstown, New Jersey: Dushkin / McGraw-Hill.

Pfeifer, Rolf / Scheier, Christian. 1999. Understanding Intelligence. Cambridge, Massachusetts: The MIT Press.

Phillips, Denis C. 1995. "The Good, the Bad, and the Ugly: The Many Faces of Constructivism." In: Educational Researcher. 24(7). October 1995. pp. 5–12.

Phillips, Robert A. 2004. "The Design Dimensions of E-Learning." In: Proceedings of the 21st Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2004). Perth, Australia. 5–8 December 2004.

Piaget, Jean. 1952. The Origins of Intelligence in Children. New York: International Universities Press.

Piaget, Jean. 1980. "The Psychogenesis of Knowledge and Its Epistemological Significance." In: Piattelli-Palmarini. (ed.). 1980. pp. 23–34.

Piattelli-Palmarini, Massimo. (ed.). 1980. Language and Learning: The Debate Between Jean Piaget and Noam Chomsky. Cambridge, Massachusetts: Harvard University Press.

Picard, Rosalind W. 1995. "Affective Computing." Technical Report 321. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. November 1995. Available Online. URL: http://affect.media.mit.edu/pdfs/95.picard.pdf. Last Access: 11 September 2009.

Picard, Rosalind W. 1997. "Does HAL Cry Digital Tears?: Emotion and Computers." In: Stork. (ed.). 1997a. pp. 279–303.

Picard, Rosalind W. 1997, 2000. Affective Computing. 1st MIT Press Paperback Edition. Cambridge, Massachusetts: The MIT Press.

Picard, Rosalind W. 2001. "Building HAL: Computers that Sense, Recognize, and Respond to Human Emotion." Technical Report 532. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. 2001. Available Online. URL: http://vismod.media.mit.edu/tech-reports/TR-532.pdf. Last Access: 20 September 2009.

Picard, Rosalind W. 2003. "What Does It Mean for a Computer to 'Have' Emotions?." In: Trappl et al. (eds.). 2003. pp. 213–235.

Picard, Rosalind W. / Papert, Seymour A. / Bender, Walter / Blumberg, Bruce / Breazeal, Cynthia / Cavallo, David / Machover, Tod / Resnick, Mitchel / Roy, Deb / Strohecker, Carol. 2004. "Affective Learning: A Manifesto." In: BT Technology Journal. 22(4). October 2004. pp. 253–269.

Picard, Rosalind W. / Vyzas, Elias / Healey, Jennifer. 2001. "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State." In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 23(10). October 2001. pp. 1175–1191.

Pierce, John R. 1961. Symbols, Signals, and Noise: The Nature and Process of Communication. New York: Harper & Brothers.

Piwek, Paul / Krenn, Brigitte / Schröder, Marc / Grice, Martine / Baumann, Stefan / Pirker, Hannes. 2002. "RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA." In: Proceedings of the Workshop "Embodied Conversational Agents – Let's Specify and Evaluate Them!." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Piwek, Paul. 2003. "An Annotated Bibliography of Affective Natural Language Generation." Version 1.3. Information Technology Research Institute (ITRI), University of Brighton. Brighton, United

Kingdom. April 2003. Available Online. URL: http://www.itri.bton.ac.uk/projects/neca/affect-bib.pdf. Last Access: 11 September 2009.

Poggi, Isabella / Pelachaud, Catherine. 2002. “Signals and Meanings of Gaze in Animated Faces.” In: McKevitt et al. (eds.). 2002. pp. 133–144.

Poggi, Isabella / Pelachaud, Catherine / de Rosis, Fiorella. 2000. “Eye Communication in a Conversational 3D Synthetic Agent.” In: AI Communications. 13(3). 2000. pp. 169–181.

Polson, Martha C. / Richardson, J. Jeffrey. (eds.). 1988. Foundations of Intelligent Tutoring Systems. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Popper, Karl R. 1972, 1979. Objective Knowledge: An Evolutionary Approach (Revised Edition). Oxford: Oxford University Press.

Portele, Thomas. 1997. “Künstliche Sprache aus dem Baukasten: Datenbasierte Sprachsynthese.” In: Spektrum der Wissenschaft, Dossier: “Kopf oder Computer.” Dossier 1997(4). October 1997. pp. 59–61.

Potamianos, Gerasimos / Neti, Chalapathy / Gravier, Guillaume / Garg, Ashutosh / Senior, Andrew W. 2003. “Recent Advances in the Automatic Recognition of Audio-Visual Speech.” In: Proceedings of the IEEE. 91(9). September 2003. pp. 1306–1326.

Potamianos, Gerasimos / Neti, Chalapathy / Luettin, Juergen / Matthews, Iain. 2009. “Audio-Visual Automatic Speech Recognition: An Overview.” To Appear in: Vatikiotis-Bateson et al. (eds.). 2009. Available Online. URL: http://www.research.ibm.com/AVSTG/MITBOOK04_REVIEW.pdf. Last Access: 11 September 2009.

Potter, Stephen. 2001. “A Survey of Knowledge Acquisition from Natural Language.” Technological Maturity Assessment (TMA). Advanced Knowledge Technologies Project. Advanced Knowledge Technologies Project, Work Package 1, AKT Challenges, Subpackage 1.1, Knowledge Acquisition, Task 1.1.2, KA and Natural Language. Division of Informatics, University of Edinburgh. Edinburgh, United Kingdom. 10 May 2001. Available Online. URL: http://www.aiai.ed.ac.uk/project/akt/work/stephenp/TMA%20of%20KAfromNL.pdf. Last Access: 11 September 2009.

Powell, Richard A. / Single, Helen M. 1996. “Focus Groups.” In: International Journal for Quality in Health Care. 8(5). 1996. pp. 499–504.

Preece, Jenny / Rogers, Yvonne / Sharp, Helen. 2002. Interaction Design: Beyond Human-Computer Interaction. New York: John Wiley & Sons.

Prendinger, Helmut / Ishizuka, Mitsuru. (eds.). 2004a. Life-Like Characters: Tools, Affective Functions, and Applications. Berlin: Springer Verlag.

Prendinger, Helmut / Ishizuka, Mitsuru. 2004b. “Introducing the Cast for Social Computing: Life-Like Characters.” In: Prendinger & Ishizuka. 2004a. pp. 3–16.

Prendinger, Helmut / Ma, Chunling / Ishizuka, Mitsuru. 2007. “Eye Movements as Indices for the Utility of Life-Like Interface Agents: A Pilot Study.” In: Interacting with Computers. 19(2). March 2007. pp. 281–292.

Prensky, Marc. 2001a. “Digital Natives, Digital Immigrants, Part 1.” In: On the Horizon. 9(5). October 2001. pp. 1–6.

Prensky, Marc. 2001b. “Digital Natives, Digital Immigrants, Part 2: Do They Really Think Differently?.” In: On the Horizon. 9(6). December 2001. pp. 1–6.

Price, Patti. 1997. “Spoken Language Understanding.” In: Cole et al. (eds.). 1997. pp. 42–49.

Proctor, Robert W. / Vu, Kim-Phuong L. (eds.). 2005. The Handbook of Human Factors in Web Design. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Purushotma, Ravi. 2005. "Commentary: You're Not Studying, You're Just…" In: Language Learning & Technology. 9(1). January 2005. pp. 80–96. Available Online. URL: http://llt.msu.edu/vol9num1/pdf/purushotma.pdf. Last Access: 11 September 2009.

Pustejovsky, James. 2001. "Computational Lexicons." In: Wilson & Keil. (eds.). 2001. pp. 160–162.

PytlikZillig, Lisa M. / Bodvarsson, Mary / Bruning, Roger. (eds.). 2005. Technology-Based Education: Bringing Researchers and Practitioners Together. Greenwich, Connecticut: Information Age Publishing.

Qvarfordt, Pernilla / Jönsson, Arne / Dahlbäck, Nils. 2003. "The Role of Spoken Feedback in Experiencing Multimodal Interfaces as Human-Like." In: Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI 2003). Vancouver, British Columbia, Canada. 5–7 November 2003.

Rabiner, Lawrence / Juang, Biing-Hwang. 1993. Fundamentals of Speech Recognition. Upper Saddle River, New Jersey: Prentice Hall.

Rafaeli, Sheizaf. 1988. "Interactivity: From New Media to Communication." In: Hawkins et al. (eds.). 1988. pp. 110–134.

Rafaeli, Sheizaf / Sudweeks, Fay. 1997. "Networked Interactivity." In: Journal of Computer-Mediated Communication. 2(4). March 1997. Available Online. URL: http://jcmc.indiana.edu/vol2/issue4/rafaeli.sudweeks.html. Last Access: 11 September 2009.

Raybould, Barry. 1995. "Performance Support Engineering: An Emerging Development Methodology for Enabling Organizational Learning." In: Performance Improvement Quarterly. 8(1). March 1995. pp. 7–22.

Reeves, Byron. 2002. "The Benefits of Interactive Online Characters." Center for the Study of Language and Information (CSLI), Stanford University. Stanford, California, USA. Available Online. URL: http://www.getabby.com/docs/Stanford_Avatar_Study.pdf. Last Access: 11 September 2009.

Reeves, Byron / Nass, Clifford I. 1996, 1998. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. 1st Paperback Edition. Stanford, California: CSLI Publications.

Reeves, Leah M. / Lai, Jennifer / Larson, James A. / Oviatt, Sharon L. / Balaji, T. S. / Buisine, Stéphanie / Collings, Penny / Cohen, Philip / Kraal, Ben / Martin, Jean-Claude / McTear, Michael F. / Raman, T. V. / Stanney, Kay M. / Su, Hui / Wang, Qian Ying. 2004. "Guidelines for Multimodal User Interface Design." In: Communications of the ACM. 47(1). January 2004. pp. 57–59.

Reiter, Ehud / Dale, Robert. 1997. "Building Applied Natural Language Generation Systems." In: Natural Language Engineering. 3(1). March 1997. pp. 57–87.

Reiter, Ehud / Dale, Robert. 2000. Building Natural Language Generation Systems. Cambridge: Cambridge University Press.

Reiter, Ehud / Sripada, Somayajulu G. / Robertson, Roma. 2003. "Acquiring Correct Knowledge for Natural Language Generation." In: Journal of Artificial Intelligence Research. 18. January–June 2003. pp. 491–516.

Renkl, Alexander. 2005. "The Worked-Out Examples Principle in Multimedia Learning." In: Mayer. (ed.). 2005a. pp. 229–245.

Rheingold, Howard. 1991. Virtual Reality. New York: Summit Books.

Rhodes, Dent M. / Azbell, Janet W. 1985. “Designing Interactive Video Instruction Professionally.” In: Training and Development Journal. 39(12). December 1985. pp. 31–33.

Rice, Ronald E. (ed.). 1984a. The New Media: Communication, Research, and Technology. Beverly Hills, California: SAGE Publications.

Rice, Ronald E. 1984b. “New Media Technology: Growth and Integration.” In: Rice. (ed.). 1984a. pp. 33–54.

Rich, Elaine. 1989. “Stereotypes and User Modeling.” In: Kobsa & Wahlster. (eds.). 1989. pp. 35–51.

Richardson, Matthew / Agrawal, Rakesh / Domingos, Pedro. 2003. “Trust Management for the Semantic Web.” In: Proceedings of the 2nd International Semantic Web Conference (ISWC 2003). Sanibel Island, Florida, USA. 20–23 October 2003.

Rickel, Jeff W. 1989. “Intelligent Computer-Aided Instruction: A Survey Organized around System Components.” In: IEEE Transactions on Systems, Man, and Cybernetics. 19(1). January–February 1989. pp. 40–57.

Rickel, Jeff W. / Johnson, W. Lewis. 1997. “Integrating Pedagogical Capabilities in a Pedagogical Agent.” In: Proceedings of the 1st International Conference on Autonomous Agents (Agents ’97). Marina del Rey, California, USA. 5–8 February 1997.

Rickel, Jeff W. / Johnson, W. Lewis. 1999. “Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control.” In: Applied Artificial Intelligence. 13(4–5). May 1999. pp. 343–382.

Rickel, Jeff W. / Johnson, W. Lewis. 2000. “Task-Oriented Collaboration with Embodied Agents in Virtual Worlds.” In: Cassell et al. (eds.). 2000a. pp. 95–122.

Rickel, Jeff W. / Johnson, W. Lewis. 2003. “Extending Virtual Humans to Support Team Training in Virtual Reality.” In: Lakemeyer & Nebel. (eds.). 2003. pp. 217–238.

Rickel, Jeff W. / Johnson, W. Lewis / Lester, James C. 1999. “Why Use an Animated Pedagogical Agent?.” In: Proceedings of the Workshop on Animated and Personified Pedagogical Agents. 9th International Conference on Artificial Intelligence in Education (AI-ED ’99). Le Mans, France. 19–23 July 1999.

Rickel, Jeff W. / Marsella, Stacy / Gratch, Jonathan / Hill, Randall W., Jr. / Traum, David R. / Swartout, William R. 2002. “Toward a New Generation of Virtual Humans for Interactive Experiences.” In: IEEE Intelligent Systems. 17(4). July–August 2002. pp. 32–38.

Rickenberg, Raoul / Reeves, Byron. 2000. “The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces.” In: CHI Letters. 2(1). April 2000. pp. 49–56. Proceedings of the Conference on Human Factors in Computing Systems (CHI 2000). The Hague, The Netherlands. 1–6 April 2000.

Riskin, Jessica. (ed.). 2007. Genesis Redux: Essays in the History and Philosophy of Artificial Life. Chicago: University of Chicago Press.

Rist, Roger / Hewer, Sue. 1996. “What Is Learning Technology?: Some Definitions.” In: Stoner. (ed.). 1996. pp. 3–5.

Rist, Thomas. 2004. “Some Issues in the Design of Character Scripting and Specification Languages: A Personal View.” In: Prendinger & Ishizuka. (eds.). 2004a. pp. 463–468.

Rist, Thomas / André, Elisabeth / Baldes, Stephan. 2003. “A Flexible Platform for Building Applications with Life-Like Characters.” In: Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI 2003). Miami, Florida, USA. 12–15 January 2003.

Rist, Thomas / André, Elisabeth / Baldes, Stephan / Gebhard, Patrick / Klesen, Martin / Kipp, Michael / Rist, Peter / Schmitt, Markus. 2004. "A Review of the Development of Embodied Presentation Agents and Their Application Fields." In: Prendinger & Ishizuka. (eds.). 2004a. pp. 377–404.

Ritchie, Graeme. 1998. "Prospects for Computational Humour." In: Proceedings of the 7th IEEE International Workshop on Robot and Human Communication (IEEE ROMAN '98). Takamatsu, Japan. 30 September–2 October 1998.

Ritchie, Graeme. 2001. "Current Directions in Computational Humour." In: Artificial Intelligence Review. 16(2). October 2001. pp. 119–135.

Robinson, John P. / Shaver, Phillip R. / Wright, Lawrence S. (eds.). 1991. Measures of Personality and Social Psychological Attitudes. San Diego: Academic Press.

Roblyer, Margaret D. / Wiencke, William R. 2004. "Exploring the Interaction Equation: Validating a Rubric to Assess and Encourage Interaction in Distance Courses." In: Journal of Asynchronous Learning Networks. 8(4). December 2004. pp. 25–37. Available Online. URL: http://www.sloan-c.org/publications/jaln/v8n4/pdf/v8n4_roblyer.pdf. Last Access: 11 September 2009.

Rogers, Everett M. 1986. Communication Technology: The New Media in Society. New York: The Free Press.

Rogers, Everett M. 1995. Diffusion of Innovations (4th ed.). New York: The Free Press.

Rogoff, Barbara / Lave, Jean. (eds.). 1984. Everyday Cognition: Its Development in Social Context. Cambridge, Massachusetts: Harvard University Press.

Rolls, Edmund T. 2003. "A Theory of Emotion, Its Functions, and Its Adaptive Value." In: Trappl et al. (eds.). 2003. pp. 11–34.

Romano, Daniela M. 2007. "The Look, the Emotion, the Language and the Behaviour of a Companion at Real-Time." In: Proceedings of the Forum "Artificial Companions in Society: Perspectives on the Present and Future." Oxford Internet Institute, University of Oxford. Oxford, United Kingdom. 26 October 2007. Available Online. URL: http://www.companions-project.org/downloads/Companions_Position_Papers_20071026.pdf. Last Access: 11 September 2009.

Romiszowski, Alexander J. 2004. "How's the E-Learning Baby?: Factors Leading to Success or Failure of an Educational Technology Innovation." In: Educational Technology. 44(1). January–February 2004. pp. 5–27.

Romme, Georges. 2002. "The Educational Value of Microworld Simulation." In: Proceedings of the 20th System Dynamics Conference. Palermo, Italy. 28 July–1 August 2002.

Rose, Ellen. 1999. "Deconstructing Interactivity in Educational Computing." In: Educational Technology. 39(1). January–February 1999. pp. 43–49.

Rosenberg, Marc J. 2001. E-Learning: Strategies for Delivering Knowledge in the Digital Age. New York: McGraw-Hill.

Rosenberg-Kima, Rinat B. / Baylor, Amy L. / Plant, E. Ashby / Doerr, Celeste E. 2007. "The Importance of Interface Agent Visual Presence: Voice Alone Is Less Effective in Impacting Young Women's Attitudes Toward Engineering." In: Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE 2007). Palo Alto, California, USA. 26–27 April 2007.

Rosenberg-Kima, Rinat B. / Baylor, Amy L. / Plant, E. Ashby / Doerr, Celeste E. 2008. "Interface Agents as Social Models for Female Students: The Effects of Agent Visual Presence and Appearance on Female Students' Attitudes and Beliefs." In: Computers in Human Behavior. 24(6). 17 September 2008. pp. 2741–2756.

Rosenschein, Stanley J. 2001. "Intelligent Agent Architecture." In: Wilson & Keil. (eds.). 2001. pp. 411–412.

Rosenthal, Robert / Jacobson, Lenore. 1992. Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development. New York: Irvington Publishers.

Rossett, Allison / Douglis, Felicia / Frazee, Rebecca V. 2003. "Strategies for Building Blended Learning." In: Learning Circuits. July 2003. Available Online. URL: http://www.astd.org/LC/2003/0703_rossett.htm. Last Access: 11 September 2009.

Rötzer, Florian. 2006. "Von Asiatischen Evas und Robotern als Zeitarbeitern." In: Telepolis. 30 November 2006. Available Online. URL: http://www.heise.de/tp/r4/artikel/24/24060/1.html. Last Access: 11 September 2009.

Rothrock, Ling / Koubek, Richard / Fuchs, Frederic / Haas, Michael / Salvendy, Gavriel. 2002. "Review and Reappraisal of Adaptive Interfaces: Toward Biologically Inspired Paradigms." In: Theoretical Issues in Ergonomics Science. 3(1). January 2002. pp. 47–84.

Rourke, Liam. 2000. "Operationalizing Social Interaction in Computer Conferencing." In: Proceedings of the 16th Annual Conference of the Canadian Association for Distance Education. Quebec City, Quebec, Canada. 3–6 May 2000.

Rourke, Liam / Anderson, Terry D. / Garrison, D. Randy / Archer, Walter. 1999. "Assessing Social Presence in Asynchronous Text-Based Computer Conferencing." In: Journal of Distance Education. 14(2). Fall 1999. pp. 50–71.

Rovai, Alfred P. / Jordan, Hope M. 2004. "Blended Learning and Sense of Community: A Comparative Analysis with Traditional and Fully Online Graduate Courses." In: The International Review of Research in Open and Distance Learning. 5(2). August 2004. Available Online. URL: http://www.irrodl.org/index.php/irrodl/article/view/192/274. Last Access: 11 September 2009.

Rowe, Jonathan P. / Ha, Eun Young. / Lester, James C. 2008. "Archetype-Driven Character Dialogue Generation for Interactive Narrative." In: Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA 2008). Tokyo, Japan. 1–3 September 2008.

Royce, Winston W. 1970. "Managing the Development of Large Software Systems: Concepts and Techniques." In: Proceedings of IEEE WESCON. Los Angeles, California, USA. August 1970.

Rudnicky, Alexander I. / Hauptmann, Alexander G. / Lee, Kai-Fu. 1994. "Survey of Current Speech Technology." In: Communications of the ACM. 37(3). March 1994. pp. 52–57.

Ruggeri, Fabrizio / Kenett, Ron S. / Faltin, Frederick W. (eds.). 2007. Encyclopedia of Statistics in Quality and Reliability. 4 Volumes. Chichester: John Wiley & Sons.

Russell, James A. / Bachorowski, Jo-Anne / Fernández-Dols, José-Miguel. 2003. "Facial and Vocal Expressions of Emotion." In: Annual Review of Psychology. 54. February 2003. pp. 329–349.

Russell, James A. / Fernández-Dols, José-Miguel. (eds.). 1997. The Psychology of Facial Expression. Cambridge: Cambridge University Press.

Russell, Rory S. 2002. "Language Use, Personality, and True Conversational Interfaces." Honors Project Report. School of Informatics, University of Edinburgh. Edinburgh, United Kingdom. June 2002. Available Online. URL: http://www.geocities.com/rorysr2002/report.pdf. Last Access: 11 September 2009.

Russell, Stuart J. / Norvig, Peter. 2003. Artificial Intelligence: A Modern Approach (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Rutter, Derek R. 1984. Looking and Seeing: The Role of Visual Communication in Social Interaction. Chichester: John Wiley & Sons.

Ruttkay, Zsófia / André, Elisabeth / Johnson, W. Lewis / Pelachaud, Catherine. (eds.). 2006. 04121 Abstracts Collection: Evaluating Embodied Conversational Agents. Dagstuhl Seminar 04121 on Evaluating Embodied Conversational Agents. Castle Dagstuhl, Wadern, Germany. 15–19 March 2004. Available Online. URL: http://drops.dagstuhl.de/opus/volltexte/2006/501/. Last Access: 5 September 2009.

Ruttkay, Zsófia / Dormann, Claire / Noot, Han. 2002. "Evaluating ECAs – What and How?." In: Proceedings of the Workshop "Embodied Conversational Agents – Let's Specify and Evaluate Them!." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Ruttkay, Zsófia / Dormann, Claire / Noot, Han. 2004. "Embodied Conversational Agents on a Common Ground: A Framework for Design and Evaluation." In: Ruttkay & Pelachaud. (eds.). 2004. pp. 27–66.

Ruttkay, Zsófia / Pelachaud, Catherine. (eds.). 2004. From Brows to Trust: Evaluating Embodied Conversational Agents. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ryokai, Kimiko / Vaucelle, Catherine / Cassell, Justine. 2003. "Virtual Peers as Partners in Storytelling and Literacy Learning." In: Journal of Computer Assisted Learning. 19(2). June 2003. pp. 195–208.

Ryu, Jeehon / Baylor, Amy L. 2005. "The Psychometric Structure of Pedagogical Agent Persona." In: Technology, Instruction, Cognition and Learning (TICL). 2(4). 2005. pp. 291–315.

Sacks, Harvey. 1995. Lectures on Conversation. Oxford: Blackwell Publishing.

Salovey, Peter / Mayer, John D. 1990. "Emotional Intelligence." In: Imagination, Cognition and Personality. 9(3). 1989–1990. pp. 185–211.

Salvaggio, Jerry / Bryant, Jennings. (eds.). 1989. Media Use in the Information Age: Emerging Patterns of Adoption and Consumer Use. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sanders, Gregory A. / Scholtz, Jean. 2000. "Measurement and Evaluation of Embodied Conversational Agents." Expanded and Re-Edited Version of Cassell et al. (eds.). 2000a. pp. 346–373. Available Online. URL: http://www.nist.gov/speech/publications/storage_paper/sanders_and_scholtz_expanded .pdf. Last Access: 11 September 2009.

Sandholm, Tuomas. 1999. "Automated Negotiation: The Best Terms for All Concerned." In: Communications of the ACM. 42(3). March 1999. pp. 84–85.

Sansone, Carol / Harackiewicz, Judith M. (eds.). 2000. Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance. San Diego: Academic Press.

Sarawagi, Sunita. 2008. "Information Extraction." In: Foundations and Trends in Databases. 1(3). 2008. pp. 261–377.

Sarbin, Theodore R. (ed.). 1986. Narrative Psychology: The Storied Nature of Human Conduct. New York: Praeger Publishers.

Sarter, Nadine B. 2006. "Multimodal Information Presentation: Design Guidance and Research Challenges." In: International Journal of Industrial Ergonomics. 36(5). May 2006. 439–445.

Saygin, Ayse P. / Cicekli, Ilyas / Akman, Varol. 2000. "Turing Test: 50 Years Later." In: Minds and Machines. 10(4). November 2000. pp. 463–518.

Schank, Roger C. 1972. "Conceptual Dependency: A Theory of Natural Language Understanding." In: Cognitive Psychology. 3(4). October 1972. pp. 552–631.

Schank, Roger C. / Abelson, Robert P. 1977. Scripts, Plans, Goals and Understanding. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schank, Roger C. / Carbonell, Jaime G., Jr. 1978. “Re: The Gettysburg Address: Representing Social and Political Acts.” Research Report 127. Department of Computer Science, Yale University. New Haven, Connecticut, USA. January 1978.

Schank, Roger C. / Neaman, Adam. 2001. “Motivation and Failure in Educational Simulation Design.” In: Forbus & Feltovich. (eds.). 2001. pp. 37–69.

Schaumburg, Heike. 2001. “Computers as Tools or as Social Actors?: The Users’ Perspective on Anthropomorphic Agents.” In: International Journal of Cooperative Information Systems. 10(1–2). March and June 2001. pp. 217–234.

Scherer, Klaus R. 1981. “Speech and Emotional States.” In: Darby. (ed.). 1981. pp. 189–220.

Scherer, Klaus R. 1999. “Appraisal Theory.” In: Dalgleish & Power. (eds.). 1999. pp. 637–663.

Scherer, Klaus R. / Ekman, Paul. (eds.). 1984. Approaches to Emotion. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Scherer, Klaus R. / Schorr, Angela / Johnstone, Tom. (eds.). 2001. Appraisal Processes in Emotion. New York: Oxford University Press.

Schiaffino, Silvia / Amandi, Analía. 2004. “User-Interface Agent Interaction: Personalization Issues.” In: International Journal of Human-Computer Studies. 60(1). January 2004. pp. 129–148.

Schmidt, Albrecht. 2002. “Ubiquitous Computing – Computing in Context.” Ph.D. Thesis. Computing Department, Lancaster University. Lancaster, United Kingdom. November 2002. Available Online. URL: http://www.comp.lancs.ac.uk/~albrecht/phd/Albrecht_Schmidt_PhD-Thesis_Ubiquitous-Computing_print1.pdf. Last Access: 11 September 2009.

Schmitz, Ulrich. 1992. Computerlinguistik: Eine Einführung. Opladen: Westdeutscher Verlag.

Schneider, Barbara / McDonald, Sarah-Kathryn. (eds.). 2007. Scale-Up in Education. Volume 2. Lanham, Maryland: Rowman & Littlefield Publishers.

Schröder, Marc. 2004. “Speech and Emotion Research: An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis.” Ph.D. Thesis. Institute of Phonetics, Saarland University. Saarbrucken, Germany. Research Report PHONUS 7. Available Online. URL: http://www.dfki.de/~schroed/articles/schroeder_phd_2004.pdf. Last Access: 11 September 2009.

Schuler, Douglas / Namioka, Aki. (eds.). 1993. Participatory Design: Principles and Practices. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schulman, Daniel / Sharma, Mayur / Bickmore, Timothy W. 2008. “The Identification of Users by Relational Agents.” In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2008). Estoril, Portugal. 12–16 May 2008.

Schulmeister, Rolf. 1997. Grundlagen Hypermedialer Lernsysteme: Theorie – Didaktik – Design (2nd ed.). Munich: Oldenbourg Verlag.

Schulmeister, Rolf. 2003. “Taxonomy of Multimedia Component Interactivity: A Contribution to the Current Metadata Debate.” In: Studies in Communication Sciences (Studi di Scienze della Communicazione). 3(3). March 2003. pp. 61–80.

Schulmeister, Rolf. 2005. “Interaktivität in Multimedia-Anwendungen.” E-teaching.org, Portalbereich Didaktisches Design / Mediengestaltung / Interaktivität. 8 November 2005. Available Online. URL:

http://www.e-teaching.org/didaktik/gestaltung/interaktiv/InteraktivitaetSchulmeister.pdf. Last Access: 11 September 2009.

Schulmeister, Rolf. 2006. eLearning: Einsichten und Aussichten. Munich: Oldenbourg Verlag.

Schulmeister, Rolf. 2008. "Gibt es eine 'Net Generation'?." Work in Progress. Center for Academic and Continuing Education, University of Hamburg. Hamburg, Germany. Available Online. URL: http://www.zhw.uni-hamburg.de/pdfs/Schulmeister_Netzgeneration.pdf. Last Access: 11 September 2009.

Schultz, Tanjev. 1999. "Interactive Options in Online Journalism: A Content Analysis of 100 U.S. Newspapers." In: Journal of Computer-Mediated Communication. 5(1). September 1999. Available Online. URL: http://jcmc.indiana.edu/vol5/issue1/schultz.html. Last Access: 11 September 2009.

Schumaker, Robert P. / Liu, Ying / Ginsburg, Mark / Chen, Hsinchun. 2006. "Evaluating Mass Knowledge Acquisition Using the ALICE Chatterbot: The AZ-ALICE Dialog System." In: International Journal of Human-Computer Studies. 64(11). November 2006. pp. 1132–1140.

Schumann, David W. / Artis, Andy / Rivera, Rachael. 2001. "The Future of Interactive Advertising Viewed Through an IMC Lens." In: Journal of Interactive Advertising. 1(2). Spring 2001. pp. 16–32. Available Online. URL: http://jiad.org/article12. Last Access: 11 September 2009.

Schunk, Dale H. 1991. Learning Theories: An Educational Perspective. New York: Macmillan.

Scriven, Michael. 1991. Evaluation Thesaurus (4th ed.). Newbury Park, California: SAGE Publications.

Searle, John R. 1979. Expression and Meaning: Studies in the Theory of Speech Acts. Cambridge: Cambridge University Press.

Seidel, John V. 1998. "Qualitative Data Analysis." Qualis Research. Colorado Springs, Colorado, USA. Available Online. URL: ftp://ftp.qualisresearch.com/pub/qda.pdf. Last Access: 11 September 2009.

Self, John A. 1988. "Student Models: What Use Are They?." In: Ercoli & Lewis. (eds.). 1988. pp. 73–86.

Self, John A. 1990. "Bypassing the Intractable Problem of Student Modelling." In: Frasson & Gauthier. (eds.). 1990. pp. 107–123.

Self, John A. 1994a. "Formal Approaches to Student Modelling." In: Greer & McCalla. (eds.). 1994. pp. 295–352.

Self, John A. 1994b. "The Role of Student Models in Learning Environments." In: Transactions of the Institute of Electronics, Information and Communication Engineers. E77-D(1). January 1994. pp. 3–8.

Self, John A. 1999. "The Defining Characteristics of Intelligent Tutoring Systems Research: ITSs Care, Precisely." In: International Journal of Artificial Intelligence in Education. 10(3–4). 1999. pp. 350–364.

Sengers, Phoebe. 1998. "Anti-Boxology: Agent Design in Cultural Context." Ph.D. Thesis. Computer Science Department and Program in Literary and Cultural Theory, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. Technical Report CMU-CS-98-151. August 1998. Available Online. URL: http://reports-archive.adm.cs.cmu.edu/anon/1998/CMU-CS-98-151.pdf. Last Access: 11 September 2009.

Sengers, Phoebe. 1999. "Designing Comprehensible Agents." In: Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI '99). Stockholm, Sweden. 31 July–6 August 1999.

Sengers, Phoebe. 2000. "Narrative Intelligence." In: Dautenhahn. (ed.). 2000. pp. 1–26.

Sengers, Phoebe. 2004. "The Agents of McDonaldization." In: Payr & Trappl. (eds.). 2004. pp. 3–19.

Serenko, Alexander. 2006. “The Use of Interface Agents for Email Notification in Critical Incidents.” In: International Journal of Human-Computer Studies. 64(11). November 2006. pp. 1084–1098.

Servan-Schreiber, David. 1987. “From Intelligent Tutoring to Computerized Psychotherapy.” In: Proceedings of the 6th National Conference on Artificial Intelligence (AAAI ’87). Seattle, Washington, USA. 13–17 July 1987.

Sevik, Kristine. 2003. “RadioWeb: The Design, Development and Formative Evaluation of Web-Based Learning Material for Radiology.” Master’s Thesis. Department of Information Science, University of Bergen. Bergen, Norway. May 2003. Available Online. URL: http://www.ub.uib.no/elpub/2003/h/704002/Hovedoppgave.pdf. Last Access: 11 September 2009.

Sewart, David / Keegan, Desmond / Holmberg, Borje. 1988. (eds.). Distance Education: International Perspectives. London: Routledge.

Shadbolt, Nigel / Berners-Lee, Tim / Hall, Wendy. 2006. “The Semantic Web Revisited.” In: IEEE Intelligent Systems. 21(3). May–June 2006. pp. 96–101.

Shafer, Glenn. 1976. A Mathematical Theory of Evidence. Princeton: Princeton University Press.

Shannon, Claude E. 1948a. “A Mathematical Theory of Communication.” Part I. In: The Bell System Technical Journal. 27(3). July 1948. pp. 379–423.

Shannon, Claude E. 1948b. “A Mathematical Theory of Communication.” Part II. In: The Bell System Technical Journal. 27(4). October 1948. pp. 623–656.

Shannon, Claude E. / Weaver, Warren. 1949. The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois Press.

Shapiro, Amy / Niederhauser, Dale. 2004. “Learning from Hypertext: Research Issues and Findings.” In: Jonassen. (ed.). 2004. pp. 605–620.

Shapiro, Stuart C. (ed.). 1992. Encyclopedia of Artificial Intelligence (2nd ed.). 2 Volumes. New York: John Wiley & Sons.

Sharples, Mike. (ed.). 2007. “Big Issues in Mobile Learning.” Report of a Workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative. Learning Sciences Research Institute, University of Nottingham. Nottingham, United Kingdom. January 2007. Available Online. URL: http://mlearning.noe-kaleidoscope.org/public/news/KALEIDOSCOPE%20REPORT_07_Big_Issues_In_Mobile_Learning.pdf. Last Access: 11 September 2009.

Shaw, Erin / Johnson, W. Lewis / Ganeshan, Rajaram. 1999. “Pedagogical Agents on the Web.” In: Proceedings of the 3rd International Conference on Autonomous Agents (Agents ’99). Seattle, Washington, USA. 1–5 May 1999.

Shaw, Erin / LaBore, Catherine / Chiu, Yuan-Chun. 2004. “Animating 2D Digital Puppets with Limited Autonomy.” In: Proceedings of the 4th International Symposium on Smart Graphics (SG 2004). Banff, Alberta Canada. 23–25 May 2004.

Shieber, Stuart M. 1986. An Introduction to Unification-Based Approaches to Grammar. Stanford, California: CSLI Publications.

Shiri, Mohammad E. / Aïmeur, Esma / Frasson, Claude. 1998a. “Student Modelling by Case-Based Reasoning.” In: Proceedings of the 4th International Conference on Intelligent Tutoring Systems (ITS ’98). San Antonio, Texas, USA. 16–19 August 1998.

Shiri, Mohammad E. / Aïmeur, Esma / Frasson, Claude. 1998b. “SARA: A Case-Based Student Modelling System.” In: Proceedings of the 4th European Workshop on Advances in Case-Based Reasoning (EWCBR ’98). Dublin, Ireland. 23–25 September 1998.

Shiri, Mohammad E. / Aïmeur, Esma / Frasson, Claude. 1998c. "Case-Based Student Modelling: Unaccessible Solution Mode." In: Proceedings of the International Conference on New Technologies of Information and Communication in Engineering Education and in Industry (NTICF '98). Rouen, France. 18 –20 November 1998.

Shneiderman, Ben. 1995. "Looking for the Bright Side of User Interface Agents." In: Interactions. 2(1). January 1995. pp. 13–15.

Shneiderman, Ben. 1997. "Direct Manipulation Versus Agents: Paths to Predictable, Controllable, and Comprehensible Interfaces." In: Bradshaw. (ed.). 1997. pp. 97–106.

Shneiderman, Ben / Maes, Pattie. 1997. "Direct Manipulation vs. Interface Agents." In: Interactions. 4(6). November–December 1997. pp. 42–61.

Shneiderman, Ben / Plaisant, Catherine. 2004. Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th ed.). Reading, Massachusetts: Addison-Wesley.

Short, John / Williams, Ederyn / Christie, Bruce. 1976. The Social Psychology of Telecommunications. London: John Wiley & Sons.

Shute, Valerie J. / Psotka, Joseph. 1996. "Intelligent Tutoring Systems: Past, Present, and Future." In: Jonassen. (ed.). 1996. pp. 570–600.

Sidner, Candace L. 2002. "Building Spoken Language Collaborative Interface Agents." Technical Report TR2002-38. Mitsubishi Electric Research Laboratories (MERL). Cambridge, Massachusetts, USA. December 2002. Available Online. URL: http://www.merl.com/reports/docs/TR2002-38.pdf. Last Access: 11 September 2009.

Siemens, George. 2006. "Connectivism: Learning Theory or Pastime for the Self-Amused?." elearnspace. 12 November 2006. Available Online. URL: http://www.elearnspace.org/Articles/connectivism_self-amused.htm. Last Access: 11 September 2009.

Silverman, David. 2005. Doing Qualitative Research: A Practical Handbook (2nd ed.). London: SAGE Publications.

Simon, Herbert A. 1996. The Sciences of the Artificial (3rd ed.). Cambridge, Massachusetts: The MIT Press.

Simon, Herbert A. 2001. "Production Systems." In: Wilson & Keil. (eds.). 2001. pp. 676–678.

Sims, Roderick C. H. 1997a. "Interactivity: A Forgotten Art?." In: Computers in Human Behavior. 13(2). May 1997. pp. 157–180.

Sims, Roderick C. H. 1997b. "Interactive Learning as an 'Emerging' Technology: A Reassessment of Interactive and Instructional Design Strategies." In: Australian Journal of Educational Technology. 13(1). Winter 1997. pp. 68–84. Available Online. URL: http://www.ascilite.org.au/ajet/ajet13/sims.html. Last Access: 11 September 2009.

Sims, Roderick C. H. 1999. "Interactivity on Stage: Strategies for Learner-Designer Communication." In: Australian Journal of Educational Technology. 15(3). Summer 1999. pp. 257–272. Available Online. URL: http://www.ascilite.org.au/ajet/ajet15/sims.html. Last Access: 11 September 2009.

Sims, Roderick C. H. 2000. "An Interactive Conundrum: Constructs of Interactivity and Learning Theory." In: Australian Journal of Educational Technology. 16(1). Fall 2000. pp. 45–57. Available Online. URL: http://www.ascilite.org.au/ajet/ajet16/sims.html. Last Access: 11 September 2009.

Sims, Roderick C. H. 2003. "Promises of Interactivity: Aligning Learner Perceptions and Expectations with Strategies for Flexible and Online Learning." In: Distance Education. 24(1). May 2003. pp. 87–103.

Singh, Harvey. 2003. “Building Effective Blended Learning Programs.” In: Educational Technology. 43(6). November–December 2003. pp. 51–54.

Singh, Munindar P. (ed.). 2004. Practical Handbook of Internet Computing. Baton Rouge: Chapman & Hall / CRC Press.

Sison, Raymund / Shimura, Masamichi. 1998. “Student Modeling and Machine Learning.” In: International Journal of Artificial Intelligence in Education. 9(1–2). 1998. pp. 128–158.

Skantze, Gabriel. 2003. “Exploring Human Error Handling Strategies: Implications for Spoken Dialogue Systems.” In: Proceedings of the International Speech Communication Association (ISCA) Research Workshop on Error Handling in Spoken Dialogue Systems. Château-d’Oex-Vaud, Switzerland. 28–31 August 2003.

Skinner, Burrhus F. 1938. The Behavior of Organisms: An Experimental Analysis. New York: Appleton-Century.

Skinner, Burrhus F. 1968. The Technology of Teaching. New York: Appleton-Century-Crofts.

Slater, Mel / Antley, Angus / Davison, Adam / Swapp, David / Guger, Christoph / Barker, Chris / Pistrang, Nancy / Sanchez-Vives, Maria V. 2006. “A Virtual Reprise of the Stanley Milgram Obedience Experiments.” In: PLoS ONE. 1(1). e39. 20 December 2006. Available Online. URL: http://www .plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000039. Last Access: 11 September 2009.

Sleeman, Derek H. / Brown, John S. (eds.). 1982. Intelligent Tutoring Systems. New York: Academic Press.

Sleeman, Derek H. / Kelly, Anthony E. / Martinak, Rosemary / Ward, Robert D. / Moore, Joyce L. 1989. “Studies of Diagnosis and Remediation with High School Algebra Students.” In: Cognitive Science. 13(4). October–December 1989. pp. 551–568.

Sleeman, Derek H. / Ward, Robert D. / Kelly, Anthony E. / Martinak, Rosemary / Moore, Joyce L. 1991. “An Overview of Recent Studies with PIXIE.” In: Goodyear. (ed.). 1991. pp. 173–185.

Smith, George W. 1991. Computers and Human Language. New York: Oxford University Press.

Smith, Patricia L. / Ragan, Tillman J. 2005. Instructional Design (3rd ed.). Hoboken, New Jersey: John Wiley & Sons.

Snyder, Charles R. / Lopez, Shane J. (eds.). 2002. Handbook of Positive Psychology. New York: Oxford University Press.

Sohn, Dongyoung / Leckenby, John D. / Jee, Joonhyung. 2003. “The Role of Expected Interactivity in Interactive Ad Processing.” In: Proceedings of the Annual Conference of the American Academy of Advertising. Denver-Broomfield, Colorado, USA. 27–30 March 2003.

Sohn, Dongyoung / Lee, Byung-Kwan. 2005. “Dimensions of Interactivity: Differential Effects of Social and Psychological Factors.” In: Journal of Computer-Mediated Communication. 10(3). April 2005. Available Online. URL: http://jcmc.indiana.edu/vol10/issue3/sohn.html. Last Access: 11 September 2009.

Sparck Jones, Karen / Galliers, Julia Rose. 1996. Evaluating Natural Language Processing Systems. Berlin: Springer Verlag.

Spradley, James P. 1980. Participant Observation. New York: Holt, Rinehart & Winston.

Sproat, Richard W. (ed.). 1998. Multilingual Text-to-Speech Synthesis: The Bell Labs Approach. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Sproat, Richard W. / Ostendorf, Mari / Hunt, Andrew. (eds.). 1999. "The Need for Increased Speech Synthesis Research: Report of the 1998 NSF Workshop for Discussing Research Priorities and Evaluation Strategies in Speech Synthesis." Arlington, Virginia, USA. 6–7 August 1998. Available Online. URL: http://www.cslu.ogi.edu/~sproatr/newindex/report10.pdf. Last Access: 11 September 2009.

Sproull, Lee / Subramani, Mani R. / Kiesler, Sara / Walker, Janet H. / Waters, Keith. 1996. "When the Interface Is a Face." In: Human-Computer Interaction. 11(2). June 1996. pp. 97–124.

Stake, Robert E. 1995. The Art of Case Study Research. Thousand Oaks, California: SAGE Publications.

Stauffer, Karen. 1996. "Student Modeling and Web-Based Learning Systems." Athabasca University. Athabasca, Alberta, Canada. 22 March 1996. Available Online. URL: http://ccism.pc.athabascau.ca/html/students/stupage/Project/initsm.htm. Last Access: 11 September 2009.

Stede, Manfred. 1996. "Lexical Semantics and Knowledge Representation in Multilingual Sentence Generation." Ph.D. Thesis. Computer Systems Research Institute (CSRI), Department of Computer Science, University of Toronto. Toronto, Canada. Technical Report CSRI-347. June 1996. Available Online. URL: ftp://ftp.cs.toronto.edu/pub/reports/csri/347/Stede-thesis.ps. Last Access: 11 September 2009.

Steffe, Leslie P. / Gale, Jerry. (eds.). 1995. Constructivism in Education. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stephenson, Neil. 1993. Snow Crash. London: ROC / Penguin Books.

Stern, Andrew. 2003. "Creating Emotional Relationships with Virtual Characters." In: Trappl et al. (eds.). 2003. pp. 333–362.

Stern, Steven E. / Mullennix, John W. / Dyson, Corrie-Lynn / Wilson, Stephen J. 1999. "The Persuasiveness of Synthetic Speech Versus Human Speech." In: Human Factors. 41(4). December 1999. pp. 588–595.

Stern, Steven E. / Mullennix, John W. / Wilson, Stephen J. 2002. "Effects of Perceived Disability on Persuasiveness of Computer Synthesized Speech." In: Journal of Applied Psychology. 87(2). April 2002. pp. 411–417.

Steuer, Jonathan S. 1992. "Defining Virtual Reality: Dimensions Determining Telepresence." In: Journal of Communication. 42(4). Fall 1992. pp. 73–93.

Stevens, Albert L. / Collins, Allan M. / Goldin, Sally E. 1982. "Misconceptions in Students' Understanding." In: Sleeman & Brown. (eds.). 1982. pp. 13–24.

Stock, Oliviero. 1996. "Password Swordfish: Verbal Humor in the Interface." In: Proceedings of the International Workshop on Computational Humor (IWCH '96). 12th Twente Workshop on Language Technology (TWLT 12). Enschede, The Netherlands. 11–14 September 1996.

Stock, Oliviero / Strapparava, Carlo. 2006. "Automatic Production of Humorous Expressions for Catching the Attention and Remembering." In: IEEE Intelligent Systems. 21(2). March–April 2006. pp. 64–67.

Stone, Brian A. / Lester, James C. 1996. "Dynamically Sequencing an Animated Pedagogical Agent." In: Proceedings of the 13th National Conference on Artificial Intelligence (AAAI '96). Portland, Oregon, USA. 4–8 August 1996.

Stone, Carolyn / Goodyear, Peter. 1995. "Constructivismo y Diseño Instruccional: Epistemologia y Construccion del Significado (Constructivism and Instructional Design: Epistemology and the

Construction of Meaning).” In: Substratum: Temas Fundamentales en Psicologia y Educacion. 2(6). 1995. pp. 55–76.

Stone, Peter / Veloso, Manuela. 1997. “Multiagent Systems: A Survey from a Machine Learning Perspective.” Technical Report CMU-CS-97-193. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. 4 December 1997. Available Online. URL: http://reports-archive.adm.cs.cmu.edu/anon/1997/CMU-CS-97-193.ps. Last Access: 12 September 2009.

Stoner, Greg. (ed.). 1996. “Implementing Learning Technology.” Learning Technology Dissemination Initiative (LTDI), Institute for Computer-Based Learning, Heriot-Watt University. Edinburgh, United Kingdom. July 1996. Available Online. URL: http://www.icbl.hw.ac.uk/ltdi/implementing-it/implt.pdf. Last Access: 11 September 2009.

Stork, David G. (ed.). 1997a. HAL’s Legacy: 2001’s Computer as Dream and Reality. Cambridge, Massachusetts: The MIT Press.

Stork, David G. 1997b. “‘I Could See Your Lips Move:’ HAL and Speechreading.” In: Stork. (ed.). 1997a. pp. 237–261.

Stork, David G. / Hennecke, Marcus E. (eds.). 1996. Speechreading by Humans and Machines: Models, Systems, and Applications. New York: Springer Verlag.

Stone, Debbie / Jarrett, Caroline / Woodroffe, Mark / Minocha, Shailey. 2005. User Interface Design and Evaluation. San Francisco: Morgan Kaufmann Publishers.

Strack, Fritz / Martin, Leonard L. / Stepper, Sabine. 1988. “Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis.” In: Journal of Personality and Social Psychology. 54(5). May 1988. pp. 768–777.

Straubhaar, Joseph / LaRose, Robert. 1996. Communications Media in the Information Society. Belmont, California: Wadsworth Publishing.

Strauss, Anselm L. 1987. Qualitative Analysis for Social Scientists. Cambridge: Cambridge University Press.

Stromer-Galley, Jennifer. 2000. “On-Line Interaction and Why Candidates Avoid It.” In: Journal of Communication. 50(4). December 2000. pp. 111–132.

Stromer-Galley, Jennifer. 2004. “Interactivity-as-Product and Interactivity-as-Process.” In: The Information Society. 20(5). November–December 2004. pp. 391–394.

Stronks, Bas / Nijholt, Anton / van der Vet, Paul / Heylen, Dirk. 2002. “Designing for Friendship: Becoming Friends with Your ECA.” In: Proceedings of the Workshop “Embodied Conversational Agents – Let’s Specify and Evaluate Them!.” 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Studer, Rudi / Benjamins, V. Richard / Fensel, Dieter. 1998. “Knowledge Engineering: Principles and Methods.” In: Data and Knowledge Engineering. 25(1–2). March 1998. pp. 161–197.

Sundar, S. Shyam. 1994. “Is Human-Computer Interaction Social or Parasocial?.” Paper Presented at the 77th Annual Meeting of the Association for Education in Journalism and Mass Communication. Atlanta, Georgia, USA. 10–13 August 1994.

Sundar, S. Shyam. 2004. “Theorizing Interactivity’s Effects.” In: The Information Society. 20(5). November–December 2004. pp. 385–389.

Sundar, S. Shyam / Kalyanaraman, Sriram / Brown, Justin. 2003. “Explicating Web Site Interactivity: Impression Formation Effects in Political Campaign Sites.” In: Communication Research. 30(1). February 2003. pp. 30–59.

Swan, Karen. 2001. “Virtual Interactivity: Design Factors Affecting Student Satisfaction and Perceived Learning in Asynchronous Online Courses.” In: Distance Education. 22(2). 2001. pp. 306–331.

Swan, Karen. 2002. “Building Learning Communities in Online Courses: The Importance of Interaction.” In: Education, Communication and Information. 2(1). May 2002. pp. 23–49.

Swartout, William R. / Gratch, Jonathan / Hill, Randall W., Jr. / Hovy, Eduard H. / Marsella, Stacy / Rickel, Jeff W. / Traum, David R. 2006. “Toward Virtual Humans.” In: AI Magazine. 27(2). Summer 2006. pp. 96–108.

Swartz, Luke H. 2003. “Why People Hate the Paperclip: Labels, Appearance, Behavior, and Social Responses to User Interface Agents.” B.S. Honors Thesis. Symbolic Systems Program, Stanford University. Stanford, California, USA. 12 June 2003. Available Online. URL: http://xenon.stanford.edu/~lswartz/paperclip/paperclip.pdf. Last Access: 12 September 2009.

Sweller, John. 2005a. “Implications of Cognitive Load Theory for Multimedia Learning.” In: Mayer. (ed.). 2005a. pp. 19–30.

Sweller, John. 2005b. “The Redundancy Principle in Multimedia Learning.” In: Mayer. (ed.). 2005a. pp. 159–167.

Sycara, Katia P. 1998. “Multiagent Systems.” In: AI Magazine. 19(2). Summer 1998. pp. 79–92.

Sycara, Katia P. / Pannu, Anandeep S. / Williamson, Mike / Zeng, Dajun / Decker, Keith. 1996. “Distributed Intelligent Agents.” In: IEEE Expert. 11(6). December 1996. pp. 36–46.

Tadepalli, Prasad / Natarajan, Balas K. 1996. “A Formal Framework for Speedup Learning from Problems and Solutions.” In: Journal of Artificial Intelligence Research. 4. January–June 1996. pp. 445–475.

Takeuchi, Yugo / Katagiri, Yasuhiro / Nass, Clifford I. / Fogg, Brian J. 1998. “Social Response and Cultural Dependency in Human-Computer Interaction.” In: Proceedings of the 5th Pacific Rim International Conference on Artificial Intelligence (PRICAI ’98). Singapore. 22–27 November 1998.

Takeuchi, Yugo / Katagiri, Yasuhiro / Nass, Clifford I. / Fogg, Brian J. 2000. “A Cultural Perspective in Social Interface.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2000). The Hague, The Netherlands. 1–6 April 2000.

Tannenbaum, Robert S. 1998. Theoretical Foundations of Multimedia. New York: W. H. Freeman.

Tapscott, Don. 1998. Growing up Digital: The Rise of the Net Generation. New York: McGraw-Hill.

Tartaro, Andrea / Cassell, Justine. 2008. “Playing with Virtual Peers: Bootstrapping Contingent Discourse in Children with Autism.” In: Proceedings of the 8th International Conference of the Learning Sciences (ICLS 2008). Utrecht, The Netherlands. 23–28 June 2008.

Tatai, Gábor / Csordás, Annamária / Szaló, Attila / Laufer, László. 2003. “The Chatbot Feeling: Towards Animated Emotional ECAs.” In: Proceedings of the 11th Portuguese Conference on Artificial Intelligence (EPIA 2003). Beja, Portugal. 4–7 December 2003.

Tate, Austin. 2001. “Planning.” In: Wilson & Keil. (eds.). 2001. pp. 652–653.

Tatham, Mark / Morton, Katherine. 2005. Developments in Speech Synthesis. Chichester: John Wiley & Sons.

Taylor, Martin M. / Néel, Françoise / Bouwhuis, Don G. (eds.). 2000. The Structure of Multimodal Dialogue II. Amsterdam: John Benjamins.

Taylor, Paul A. 2000. "Concept-to-Speech Synthesis by Phonological Structure Matching." In: Philosophical Transactions of the Royal Society, Series A. 358(1769). 15 April 2000. pp. 1403–1417.

Tebelskis, Joe. 1995. "Speech Recognition Using Neural Networks." Ph.D. Thesis. School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA. Technical Report CMU-CS-95-142. May 1995. Available Online. URL: http://reports-archive.adm.cs.cmu.edu/anon/1995/CMU-CS-95-142.ps. Last Access: 12 September 2009.

Ten Have, Paul. 1999. Doing Conversation Analysis: A Practical Guide. London: SAGE Publications.

Tergan, Sigmar O. 1997. "Multiple Views, Contexts, and Symbol Systems in Learning with Hypertext / Hypermedia: A Critical Review of Research." In: Educational Technology. 37(4). July–August 1997. pp. 5–18.

Theune, Mariët. 2003. "Natural Language Generation for Dialogue: System Survey." Technical Report TR-CTIT-03-22. Center for Telematics and Information Technology (CTIT), University of Twente. Enschede, The Netherlands. 6 May 2003. Available Online. URL: http://www.ub.utwente.nl/webdocs/ctit/1/000000c5.pdf. Last Access: 12 September 2009.

Thomas, Frank / Johnston, Ollie. 1981, 1984. Disney Animation: The Illusion of Life (2nd pr.). New York: Abbeville Press.

Thurlow, Crispin / Lengel, Laura / Tomic, Alice. 2004. Computer Mediated Communication. Thousand Oaks, California: SAGE Publications.

Thweatt, Katherine S. / McCroskey, James C. 1996. "Teacher Nonimmediacy and Misbehavior: Unintentional Negative Communication." In: Communication Research Reports. 13(2). Fall 1996. pp. 198–204.

Tian, Ying-Li / Kanade, Takeo / Cohn, Jeffrey F. 2005. "Facial Expression Analysis." In: Li & Jain. (eds.). 2005. pp. 247–275.

Timmerman, Annemarie. 2002. "Introduction to the Application of Web-Based Surveys." ERIC Clearinghouse on Information and Technology. Syracuse, New York, USA. ERIC Identifier ED474097. August 2002. Available Online. URL: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a/da/55.pdf. Last Access: 12 September 2009.

Ting-Toomey, Stella. 1999. Communicating Across Cultures. New York: Guilford Publications.

Tirassa, Maurizio. 1997. "Mental States in Communication." In: Proceedings of the 2nd European Conference on Cognitive Science (ECCS '97). Manchester, United Kingdom. 9–11 April 1997.

Toffler, Alvin. 1970. Future Shock. New York: Random House.

Tognazzini, Bruce. 1993. "Principles, Techniques, and Ethics of Stage Magic and Their Application to Human Interface Design." In: Proceedings of the International Joint Conference on Human-Computer Interaction and Human Factors in Computing Systems (INTERCHI '93). Amsterdam, The Netherlands. 24–29 April 1993. Published Online as: "Magic and Software Design." AskTog, Nielsen Norman Group. Fremont, California, USA. Available Online. URL: http://www.asktog.com/papers/magic.html. Last Access: 12 September 2009.

Torres, Obed E. / Cassell, Justine / Prevost, Scott. 1997. "Modeling Gaze Behavior as a Function of Discourse Structure." In: Proceedings of the 1st International Workshop on Human-Computer Conversation. Bellagio, Italy. 14–16 July 1997.

Trappl, Robert / Petta, Paolo. (eds.). 1997. Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents. Berlin: Springer Verlag.

Trappl, Robert / Petta, Paolo / Payr, Sabine. (eds.). 2003. Emotions in Humans and Artifacts. Cambridge, Massachusetts: The MIT Press.

Tremayne, Mark / Dunwoody, Sharon. 2001. “Interactivity, Information Processing, and Learning on the World Wide Web.” In: Science Communication. 23(2). December 2001. pp. 111–134.

Triola, Mario F. 2006. Essentials of Statistics (3rd ed.). Harlow, United Kingdom: Addison-Wesley Longman.

Trost, Harald. 2003. “Morphology.” In: Mitkov. (ed.). 2003. pp. 25–47.

Trujillo, Arturo. 1999. Translation Engines: Techniques for Machine Translation. London: Springer Verlag.

Tsiriga, Victoria / Virvou, Maria. 2003. “Modelling the Student to Individualise Tutoring in a Web-Based ICALL.” In: International Journal of Continuing Engineering Education and Life-Long Learning. 13(3–4). 2003. pp. 350–365.

Tucker, Allen B. (ed.). 2004. The Computer Science Handbook (2nd ed.). Boca Raton: Chapman & Hall / CRC Press.

Turing, Alan M. 1950. “Computing Machinery and Intelligence.” In: Mind. 59(236). October 1950. pp. 433–460.

Turk, Matthew / Robertson, George. 2000. “Perceptual User Interfaces.” In: Communications of the ACM. 43(3). March 2000. pp. 32–34.

Turk, Matthew. 2001. “Perceptual User Interfaces.” In: Earnshaw et al. (eds.). 2001. pp. 39–51.

Turkle, Sherry. 1984. The Second Self: Computers and the Human Spirit. New York: Simon & Schuster.

Turkle, Sherry. 1995. Life on the Screen: Identity in the Age of the Internet. New York: Simon & Schuster.

Turkle, Sherry. 2003. “Relational Artifacts, Children, and Elders.” Technical Report. National Science Foundation. Cambridge, Massachusetts, USA. November 2003. NSF Grant SES-0115668.

Turkle, Sherry. 2005. “Relational Artifacts / Children / Elders: The Complexities of CyberCompanions.” In: Proceedings of the Cognitive Science Society 2005 Workshop “Toward Social Mechanisms of Android Science.” Stresa, Italy. 25–26 July 2005.

Turkle, Sherry. 2006a. “Artificial Intelligence at 50: From Building Intelligence to Nurturing Sociabilities.” Paper Presented at the Dartmouth Artificial Intelligence Conference. Hanover, New Hampshire, USA. 13–15 July 2006.

Turkle, Sherry. 2006b. “A Nascent Robotics Culture: New Complicities for Companionship.” Paper Presented at the Workshop on Human Implications of Human-Robot Interaction (HRI). 21st National Conference on Artificial Intelligence (AAAI 2006). Boston, Massachusetts, USA. 16–20 July 2006.

Turkle, Sherry / Taggart, William / Kidd, Cory D. / Dasté, Olivia. 2006. “Relational Artifacts with Children and Elders: The Complexities of Cybercompanionship.” In: Connection Science. 18(4). December 2006. pp. 347–361.

TV&C. 2001. “Structured Interviews.” In: Getting Technical with TV&C. 3(2). Fall 2001. The Test Validation and Construction Unit (TV&C) of the California State Personnel Board. Available Online. URL: http://www.spb.ca.gov/WorkArea/downloadasset.aspx?id=1256. Last Access: 12 September 2009.

Tzoukermann, Evelyne / Klavans, Judith L. / Strzalkowski, Tomek. 2003. “Information Retrieval.” In: Mitkov. (ed.). 2003. pp. 529–544.

Unger, Tobias. 2006a. "The VLC Workbooks." In: Handke & Franke. (eds.). 2006. pp. 92–135.

Unger, Tobias. 2006b. "Explorative Learning on the VLC: Linguistic Fieldwork." In: Handke & Franke. (eds.). 2006. pp. 168–181.

Uschold, Mike / Gruninger, Michael. 1996. "Ontologies: Principles, Methods, and Applications." In: The Knowledge Engineering Review. 11(2). June 1996. pp. 93–136.

Uszkoreit, Hans. 2001. "Language Technology: A First Overview." Deutsches Forschungszentrum für Künstliche Intelligenz (German Research Center for Artificial Intelligence). Saarbrucken, Germany. Available Online. URL: http://www.dfki.de/~hansu/LT.pdf. Last Access: 12 September 2009.

Valli, Alessandro. 2004. "Notes on Natural Interaction." Technical Report. Natural Interaction, iO Research Center. Chianti, Tuscany, Italy. July 2004. Available Online. URL: http://www.idemployee.id.tue.nl/g.w.m.rauterberg/Movies/NotesOnNaturalInteraction.pdf. Last Access: 12 September 2009.

Valvoda, Jakob T. 2007. "Virtual Humanoids and Presence in Virtual Environments." Ph.D. Thesis. Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen University. Aachen, Germany. 14 September 2007. Available Online. URL: http://darwin.bth.rwth-aachen.de/opus3/volltexte/2007/2014/pdf/Valvoda_Jakob.pdf. Last Access: 12 September 2009.

Van Deemter, Kees / Krahmer, Emiel / Theune, Mariët. 2005. "Real Versus Template-Based Natural Language Generation: A False Opposition?." In: Computational Linguistics. 31(1). March 2005. pp. 15–24.

Van Deemter, Kees / Krenn, Brigitte / Piwek, Paul / Klesen, Martin / Schröder, Marc / Baumann, Stefan. 2008. "Fully Generated Scripted Dialogue for Embodied Agents." In: Artificial Intelligence. 172(10). June 2008. pp. 1219–1244.

Van Dijk, Jan. 1999. The Network Society: Social Aspects of New Media. London: SAGE Publications.

Van Kuppevelt, Jan / Dybkjaer, Laila / Bernsen, Niels O. (eds.). 2005. Advances in Natural Multimodal Dialogue Systems. Dordrecht, The Netherlands: Springer Verlag.

VanLehn, Kurt. 1982. "Bugs Are Not Enough: Empirical Studies of Bugs, Impasses and Repairs in Procedural Skills." In: The Journal of Mathematical Behavior. 3(2). 1982. pp. 3–71.

VanLehn, Kurt. 1988. "Student Modeling." In: Polson & Richardson. (eds.). 1988. pp. 55–78.

VanLehn, Kurt. 1990. Mind Bugs: The Origins of Procedural Misconceptions. Cambridge, Massachusetts: The MIT Press.

VanLehn, Kurt / Freedman, Reva / Jordan, Pamela / Murray, Charles / Osan, Remus / Ringenberg, Michael / Rosé, Carolyn / Schulze, Kay G. / Shelby, Robert H. / Treacy, Donald J. / Weinstein, Anders / Wintersgill, Mary C. 2000. "Fading and Deepening: The Next Step for Andes and Other Model-Tracing Tutors." In: Proceedings of the 5th International Conference on Intelligent Tutoring Systems (ITS 2000). Montreal, Quebec, Canada. 19–23 June 2000.

VanLehn, Kurt / Graesser, Arthur C. / Jackson, G. Tanner / Jordan, Pamela / Olney, Andrew / Rosé, Carolyn P. 2007. "When Are Tutorial Dialogues More Effective Than Reading?." In: Cognitive Science. 31(1). February 2007. pp. 3–62.

VanLehn, Kurt / Lynch, Collin / Schulze, Kay G. / Shapiro, Joel A. / Shelby, Robert H. / Taylor, Linwood / Treacy, Donald J. / Weinstein, Anders / Wintersgill, Mary C. 2005. "The Andes Physics Tutoring System: Lessons Learned." In: International Journal of Artificial Intelligence in Education. 15(3). 2005. pp. 147–204.

VanLehn, Kurt / Martin, Joel. 1997. "Evaluation of an Assessment System Based on Bayesian Student Modeling." In: International Journal of Artificial Intelligence in Education. 8(2). 1997. pp. 179–221.

Van Mulken, Susanne / André, Elisabeth / Müller, Jochen. 1998. “The Persona Effect: How Substantial Is It?.” In: Proceedings of the 13th British Computer Society Annual Conference on Human-Computer Interaction (HCI ’98). Sheffield, United Kingdom. 1–4 September 1998.

Vassileva, Julita / Wasson, Barbara J. 1996. “Instructional Planning Approaches: From Tutoring towards Free Learning.” In: Proceedings of the European Conference on Artificial Intelligence in Education (EuroAIED ’96). Lisbon, Portugal. 30 September – 2 October 1996.

Vatikiotis-Bateson, Eric / Bailly, Gérard / Perrier, Pascal. (eds.). 2009. Audio-Visual Speech Processing. Cambridge, Massachusetts: The MIT Press.

Veletsianos, George / Yerasimou, Theano / Doering, Aaron. 2005. “The Role of Intelligent Agents on Learner Performance.” In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2005). Montreal, Quebec, Canada. 27 June–2 July 2005.

Verschuren, Piet / Doorewaard, Hans. 1999. Designing a Research Project. Utrecht, The Netherlands: Lemma B. V.

Vinayagamoorthy, Vinoba / Steed, Anthony / Slater, Mel. 2005. “Building Characters: Lessons Drawn from Virtual Environments.” In: Proceedings of the Cognitive Science Society 2005 Workshop “Toward Social Mechanisms of Android Science.” Stresa, Italy. 25–26 July 2005.

Viola, Paul / Jones, Michael. 2001. “Rapid Object Detection Using a Boosted Cascade of Simple Features.” In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001). Kauai, Hawaii, USA. 9–14 December 2001.

Vlassis, Nikos. 2007. A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence. San Rafael, California: Morgan & Claypool Publishers.

Vossen, Piek. 2003. “Ontologies.” In: Mitkov. (ed.). 2003. pp. 464–482.

Vygotsky, Lev S. 1962. Thought and Language. Cambridge, Massachusetts: The MIT Press.

Vygotsky, Lev S. / Cole, Michael / John-Steiner, Vera / Scribner, Sylvia / Souberman, Ellen. (eds.). 1978. Mind in Society: The Development of Higher Psychological Processes. Cambridge, Massachusetts: Harvard University Press.

Wagner, Dirk N. 2000. “Software Agents Take the Internet as a Shortcut to Enter Society: A Survey of New Actors to Study for Social Theory.” In: First Monday. 5(7). July 2000. Available Online. URL: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/771/680. Last Access: 12 September 2009.

Wagner, Ellen D. 1994. “In Support of a Functional Definition of Interaction.” In: The American Journal of Distance Education. 8(2). 1994. pp. 6–29.

Wagner, Ellen D. 2006. “On Designing Interaction Experiences for the Next Generation of Blended Learning.” In: Bonk & Graham. (eds.). 2006. pp. 41–55.

Wahlster, Wolfgang. 1988. “Distinguishing User Models from Discourse Models.” In: Computational Linguistics. 14(3). September 1988. pp. 101–103.

Wahlster, Wolfgang. (ed.). 2000. Verbmobil: Foundations of Speech-to-Speech Translation. Berlin: Springer Verlag.

Wahlster, Wolfgang. 2008. “Von Suchmaschinen zu Antwortmaschinen: Semantische Technologien und Benutzerpartizipation im Web 3.0.” In: Mattern. (ed.). 2008. pp. 59–75.

Waibel, Alex / Vo, Minh Tue / Duchnowski, Paul / Manke, Stefan. 1996. “Multimodal Interfaces.” In: Artificial Intelligence Review. 10(3–4). August 1996. pp. 299–319.

Wald, Mike. 2006. “Learning Through Multimedia: Automatic Speech Recognition Enabling Accessibility and Interaction.” In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2006). Orlando, Florida, USA. 26–30 June 2006.

Walker, Janet H. / Sproull, Lee / Subramani, Mani R. 1994. “Using a Human Face in an Interface.” In: Proceedings of the Conference on Human Factors in Computing Systems (CHI ’94). Boston, Massachusetts, USA. 24–28 April 1994.

Wallace, Raven M. 2003. “Online Learning in Higher Education: A Review of Research on Interactions Among Teachers and Students.” In: Education, Communication and Information. 3(2). July 2003. pp. 241–280.

Wallace, Richard S. 2008. “The Anatomy of A.L.I.C.E.” In: Epstein et al. (eds.). 2008. pp. 181–210.

Wang, Liang / Hu, Weiming / Tan, Tienu. 2003. “Recent Developments in Human Motion Analysis.” In: Pattern Recognition. 36(3). March 2003. pp. 585–601.

Wang, Ning / Johnson, W. Lewis / Rizzo, Paola / Shaw, Erin / Mayer, Richard E. 2005. “Experimental Evaluation of Polite Interaction Tactics for Pedagogical Agents.” In: Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2005). San Diego, California, USA. 9–12 January 2005.

Wang, Ning / Johnson, W. Lewis / Mayer, Richard E. / Rizzo, Paola / Shaw, Erin / Collins, Heather. 2008. “The Politeness Effect: Pedagogical Agents and Learning Outcomes.” In: International Journal of Human-Computer Studies. 66(2). February 2008. pp. 98–112.

Wardrip-Fruin, Noah / Montfort, Nick. (eds.). 2003. The New Media Reader. Cambridge, Massachusetts: The MIT Press.

Warnick, Barbara / Xenos, Michael / Endres, Danielle / Gastil, John. 2005. “Effects of Campaign-to-User and Text-Based Interactivity in Political Candidate Campaign Web Sites.” In: Journal of Computer-Mediated Communication. 10(3). April 2005. Available Online. URL: http://jcmc.indiana .edu/vol10/issue3/warnick.html. Last Access: 12 September 2009.

Wasson, Barbara J. 1985. “Student Models: The Genetic Graph Approach.” Master’s Thesis. Department of Computer Science, University of Waterloo. Waterloo, Ontario, Canada. Research Report CS-85-10. May 1985. Available Online. URL: http://www.cs.uwaterloo.ca/research/tr/1985/CS-85-10.pdf. Last Access: 12 September 2009.

Wasson, Barbara J. / Ludvigsen, Sten / Hoppe, Ulrich. (eds.). 2003. Designing for Change in Networked Learning Environments. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Wayner, Peter. 1995. “Free Agents.” In: Byte. 20(3). March 1995. pp. 105–114.

Webb, Geoffrey I. / Pazzani, Michael J. / Billsus, Daniel. 2001. “Machine Learning for User Modeling.” In: User Modeling and User-Adapted Interaction. 11(1–2). March 2001. pp. 19–29.

Weber, Donald C. / Skillings, John H. 2000. A First Course in the Design of Experiments: A Linear Models Approach. New York: CRC Press.

Weiss, Gerhard. (ed.). 1999. Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. Cambridge, Massachusetts: The MIT Press.

Weiten, Wayne / Lloyd, Margaret A. / Dunn, Dana S. / Hammer, Elizabeth Y. 2009. Psychology Applied to Modern Life: Adjustment in the 21st Century (9th ed.). Belmont, California: Wadsworth Cengage Learning.

Weizenbaum, Joseph. 1966. "ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine." In: Communications of the ACM. 9(1). January 1966. pp. 36–45.

Weizenbaum, Joseph. 1976. Computer Power and Human Reason: From Judgment to Calculation. San Francisco: W. H. Freeman.

Weller, Herman G. 1988. "Interactivity in Microcomputer-Based Instruction: Its Essential Components and How It Can Be Enhanced." In: Educational Technology. 28(2). February 1988. pp. 23–27.

Welsh, Tom. 2003. "The Semantic Web: Proof, Trust, and Security." In: The Cutter Edge. 23 September 2003. Available Online. URL: http://www.cutter.com/research/2003/edge030923.html. Last Access: 12 September 2009.

Wenger, Etienne. 1987. Artificial Intelligence and Tutoring Systems. Los Altos, California: Morgan Kaufmann Publishers.

Wenger, Etienne. 1998. Communities of Practice: Learning, Meaning, and Identity. Cambridge: Cambridge University Press.

Wexelblat, Alan / Maes, Pattie. 1997. "Issues for Software Agent UI." Unpublished Manuscript. Software Agents Group, Media Laboratory, Massachusetts Institute of Technology (MIT). Cambridge, Massachusetts, USA. 4 April 1997. Available Online. URL: http://citeseer.ist.psu.edu/wexelblat97 issues.html. Last Access: 9 September 2009.

Wharton, Cathleen / Rieman, John / Lewis, Clayton / Polson, Peter. 1994. "The Cognitive Walkthrough Method: A Practitioner's Guide." In: Nielsen & Mack. (eds.). 1994. pp. 105–140.

White, Jim. 1996. "Mobile Agents White Paper." General Magic, Inc. Mountain View, California, USA. Available Online. URL: http://www.klynch.com/documents/agents/index.html. Last Access: 9 September 2009.

Wiener, Norbert. 1948. Cybernetics, or Control and Communication in the Animal and the Machine. New York: John Wiley & Sons.

Wiggins, Jerry S. (ed.). 1996. The Five-Factor Model of Personality: Theoretical Perspectives. New York: The Guilford Press.

Wilkinson, Johan. 2006. "Art History Goes Graphic." Curriculum Services Canada (CSC). Available Online. URL: http://www.curriculum.org/tcf/teachers/projects/repository/ArtHistory.pdf. Last Access: 9 September 2009.

Williams, Frederick / Rice, Ronald E. / Rogers, Everett M. 1988. Research Methods and the New Media. New York: The Free Press.

Wilson, Michael. 1997. "Metaphor to Personality: The Role of Animation in Intelligent Interface Agents." In: Proceedings of the Workshop "Animated Interface Agents: Making Them Intelligent." 15th International Joint Conference on Artificial Intelligence (IJCAI '97). Nagoya, Japan. 23–29 August 1997.

Wilson, Robert A. / Keil, Frank C. (eds.). 1999, 2001. The MIT Encyclopedia of the Cognitive Sciences (MITECS). 1st MIT Press Paperback Edition. Cambridge, Massachusetts: The MIT Press.

Wilson, Ruth / Villa, Robert. 2002. "Survey on Methods and Standards for Student Modelling." Version 1.3. Public Deliverable. Diogene Project. University of Strathclyde. Glasgow, United Kingdom. 10 September 2002. Available Online. URL: http://www.crmpa.it/diogene/files/sep2002/D2.2%20 Survey%20on%20Student%20Model%20Representation%20V1.3.pdf. Last Access: 9 September 2009.

Winograd, Terry. 1983. Language as a Cognitive Process, Volume 1: Syntax. Reading, Massachusetts: Addison-Wesley.

Winston, Patrick H. (ed.). 1975. The Psychology of Computer Vision. New York: McGraw-Hill.

Wise, Barbara / Cole, Ronald A. / van Vuuren, Sarel / Schwartz, Scott / Snyder, Lynn / Ngampatipatpong, Nattawut / Tuantranont, Jariya / Pellom, Bryan. 2007. "Learning to Read with a Virtual Tutor: Foundations to Literacy." In: Kinzer & Verhoeven. (eds.). 2007. pp. 31–76.

Wisler, Carl / Datta, Lois-ellin / Silberman, George / Pickett, Penny. 1992. "Quantitative Data Analysis: An Introduction." United States General Accounting Office (GAO), Program Evaluation and Methodology Division. Washington, District of Columbia, USA. Report No. GAO/PEMD-10.1.11. May 1992. Available Online. URL: http://www.gao.gov/special.pubs/ pe10111.pdf. Last Access: 7 September 2009.

Wolcott, Linda L. 1996. "Distant, But Not Distanced: A Learner-Centered Approach to Distance Education." In: TechTrends. 41(5). October 1996. pp. 23–27.

Wood, Terry. 1995. "From Alternative Epistemologies to Practice in Education: Rethinking What It Means to Teach and Learn." In: Steffe & Gale. (eds.). 1995. pp. 331–339.

Woodill, Gary. 2004. "Where Is the Learning in E-Learning?: A Critical Analysis of the E-Learning Industry." Operitel Corporation. Peterborough, Ontario, Canada. March 2004. Available Online. URL: http://www.e-learningguru.com/wpapers/e-Learning_analysis.pdf. Last Access: 12 September 2009.

Woods, Robert H., Jr. / Baker, Jason D. 2004. "Interaction and Immediacy in Online Learning." In: The International Review of Research in Open and Distance Learning. 5(2). August 2004. Available Online. URL: http://www.irrodl.org/index.php/irrodl/article/view/186/268. Last Access: 12 September 2009.

Woods, William A. 1970. "Transition Network Grammars for Natural Language Analysis." In: Communications of the ACM. 13(10). October 1970. pp. 591–606.

Wooldridge, Michael J. 1999. "Intelligent Agents." In: Weiss. (ed.). 1999. pp. 27–78.

Wooldridge, Michael J. 2002. An Introduction to MultiAgent Systems. Chichester: John Wiley & Sons.

Wooldridge, Michael J. / Jennings, Nicholas R. 1995. "Intelligent Agents: Theory and Practice." In: The Knowledge Engineering Review. 10(2). June 1995. pp. 115–152.

Wooldridge, Michael J. / Veloso, Manuela M. (eds.). 1999. Artificial Intelligence Today: Recent Trends and Developments. Berlin: Springer Verlag.

Wu, Guohua M. 1999. "Perceived Interactivity and Attitude towards Web Sites." In: Proceedings of the Annual Conference of the American Academy of Advertising. Albuquerque, New Mexico, USA. 26–29 March 1999.

Wu, Guohua M. 2000. "The Role of Perceived Interactivity in Interactive Ad Processing." Ph.D. Thesis. Department of Advertising, The University of Texas at Austin. Austin, Texas, USA. 2000.

Wu, Ying / Huang, Thomas S. 1999. "Vision-Based Gesture Recognition: A Review." In: Proceedings of the International Gesture Workshop (GW '99). Gif-sur-Yvette, France. 17–19 March 1999.

Xiao, Jun. 2006. "Empirical Studies on Embodied Conversational Agents." Ph.D. Thesis. Georgia Institute of Technology. Atlanta, Georgia, USA. December 2006. Available Online. URL: http://etd .gatech.edu/theses/available/etd-09222006-140611/unrestricted/xiao_jun_200612_phd.pdf. Last Access: 12 September 2009.

Xiao, Jun / Stasko, John / Catrambone, Richard. 2002. "Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation." In: Proceedings of the Workshop "Embodied Conversational

Agents – Let's Specify and Evaluate Them!." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002). Bologna, Italy. 15–19 July 2002.

Yacci, Michael. 2000. "Interactivity Demystified: A Structural Definition for Distance Education and Intelligent CBT." In: Educational Technology. 40(4). July–August 2000. pp. 5–16.

Yacci, Michael / Hyman, Paul. 2001. "Factors Contributing to Ideal Instructional Interactivity." In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA 2001). Tampere, Finland. 25–30 June 2001.

Yee, Nick / Bailenson, Jeremy N. / Rickertsen, Kathryn. 2007. "A Meta-Analysis of the Impact of the Inclusion and Realism of Human-Like Faces on User Experiences in Interfaces." In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2007). San Jose, California, USA. 28 April–3 May 2007.

Yin, Robert K. 2003. Case Study Research: Design and Methods (3rd ed.). Thousand Oaks, California: SAGE Publications.

Young, Lisa D. 2003. "Bridging Theory and Practice: Developing Guidelines to Facilitate the Design of Computer-Based Learning Environments." In: Canadian Journal of Learning and Technology. 29(3). Fall 2003. Available Online. URL: http://www.cjlt.ca/index.php/cjlt/article/view/90/84. Last Access: 12 September 2009.

Zadeh, Lotfi A. 1965. "Fuzzy Sets." In: Information and Control. 8(3). June 1965. pp. 338–353.

Zand, Dale E. 1972. "Trust and Managerial Problem Solving." In: Administrative Science Quarterly. 17(2). June 1972. pp. 229–239.

Zboril, Daniel. 1995. "Einführung in die Sprachsynthese." Department of Phonetics and Speech Communication, University of Munich. Munich, Germany. No Longer Available Online. URL: http://www.phonetik.uni-muenchen.de/HS/Synthese.html. Last Access: 12 August 1998.

Zebrowitz, Leslie A. 1997. Reading Faces: Window to the Soul?. Boulder, Colorado: Westview Press.

Zhang, Ping / Galletta, Dennis. (eds.). 2006. Human-Computer Interaction and Management Information Systems: Foundations. Armonk, New York: M. E. Sharpe, Inc.

Zheng, Zhihong / Pantic, Maja / Roisman, Glenn I. / Huang, Thomas S. 2009. "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions." In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 31(1). January 2009. pp. 39–58.

Zhong, Ning / Liu, Jiming / Yao, Yiyu. (eds.). 2003. Web Intelligence. Berlin: Springer Verlag.

Ziegler, Cai. 2007. "Sinn oder nicht Sinn: Vom Suchen und Finden der Semantik im Web." In: c't – Magazin für Computertechnik. 2007(21). 1 October 2007. pp. 172–179.

Zimmerman, John / Ayoob, Ellen / Forlizzi, Jodi / McQuaid, Mick. 2005. "Putting a Face on Embodied Interface Agents." In: Proceedings of the Conference on Designing Pleasurable Products and Interfaces (DPPI 2005). Eindhoven, The Netherlands. 24–27 October 2005.

Zue, Victor W. 1997. "Conversational Interfaces: Advances and Challenges." Keynote Address. In: Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech '97). Rhodes, Greece. 22–25 September 1997.

Zue, Victor W. / Cole, Ronald A. / Ward, Wayne. 1997. "Speech Recognition." In: Cole et al. (eds.). 1997. pp. 3–10.

Zukerman, Ingrid / Albrecht, David W. 2001. "Predictive Statistical Models for User Modeling." In: User Modeling and User-Adapted Interaction. 11(1–2). March 2001. pp. 5–18.

Zukerman, Ingrid / Litman, Diane. 2001. “Natural Language Processing and User Modeling: Synergies and Limitations.” In: User Modeling and User-Adapted Interaction. 11(1–2). March 2001. pp. 129–158.

Internet Resources

[INT 1] http://www.interaction-design.org/encyclopedia/artifact.html, 4 September 2009.

[INT 2] http://www.w3.org/, 4 September 2009.

[INT 3] http://www.linguistics-online.com/, 4 September 2009.

[INT 4] http://internettime.com/itimegroup/elephant.htm, 4 September 2009.

[INT 5] http://www.jaycross.com/, 4 September 2009.

[INT 6] http://wordnetweb.princeton.edu/perl/webwn?s=learning, 4 September 2009.

[INT 7] http://www.umich.edu/~ed626/define.html, 4 September 2009.

[INT 8] http://en.wikipedia.org/w/index.php?title=Learning_theory_(education)&oldid=274604842, 4 September 2009.

[INT 9] http://www.funderstanding.com/content/behaviorism, 4 September 2009.

[INT 10] http://en.wikipedia.org/w/index.php?title=Classical_conditioning&oldid=273664463, 4 September 2009.

[INT 11] http://en.wikipedia.org/w/index.php?title=Operant_conditioning&oldid=158933932, 4 September 2009.

[INT 12] http://en.wikipedia.org/w/index.php?title=Programmed_instruction&oldid=269108705, 4 September 2009.

[INT 13] http://starfsfolk.khi.is/solrunb/cognitiv.htm, 4 September 2009.

[INT 14] http://en.wikipedia.org/w/index.php?title=Zone_of_proximal_development &oldid=269036349, 4 September 2009.

[INT 15] http://en.wikipedia.org/w/index.php?title=Situated_learning&oldid=252559623, 4 September 2009.

[INT 16] http://www.ewenger.com/theory/index.htm, 4 September 2009.

[INT 17] http://en.wikipedia.org/w/index.php?title=Relativism&oldid=274851545, 4 September 2009.

[INT 18] http://de.wikipedia.org/w/index.php?title=Hypertext&oldid=57703145, 4 September 2009.

[INT 19] http://en.wikipedia.org/w/index.php?title=Hypertext&oldid=278314148, 4 September 2009.

[INT 20] http://en.wikipedia.org/w/index.php?title=Hypermedia&oldid=278374920, 4 September 2009.

[INT 21] http://en.wikipedia.org/w/index.php?title=Multimedia&oldid=277447716, 4 September 2009.

[INT 22] http://www.flickr.com/, 4 September 2009.

[INT 23] http://del.icio.us/, 4 September 2009.

[INT 24] http://www.wikipedia.org/, 4 September 2009.

[INT 25] http://en.wikipedia.org/w/index.php?title=Wikipedia:About&oldid=278665452, 4 September 2009.

[INT 26] http://en.wikipedia.org/w/index.php?title=Web_syndication&oldid=262913128, 4 September 2009.

[INT 27] http://en.wikipedia.org/w/index.php?title=Web_feed&oldid=276780370, 4 September 2009.

[INT 28] http://en.wikipedia.org/w/index.php?title=Podcast&oldid=278400627, 4 September 2009.

[INT 29] http://en.wikipedia.org/w/index.php?title=Learning_management_system &oldid=278662772, 4 September 2009.

[INT 30] http://en.wikipedia.org/w/index.php?title=Virtual_learning_environment &oldid=277186105, 4 September 2009.

[INT 31] http://en.wiktionary.org/w/index.php?title=content_management_system &oldid=4104646, 4 September 2009.

[INT 32] http://en.wikipedia.org/w/index.php?title=Web_content_management_system &oldid=278929341, 4 September 2009.

[INT 33] http://vct.online.uni-marburg.de/, 4 September 2009.

[INT 34] http://voctrainer.online.uni-marburg.de/, 4 September 2009.

[INT 35] http://en.wikipedia.org/w/index.php?title=Belief-Desire-Intention_software_model &oldid=310921939, 4 September 2009.

[INT 36] http://en.wikipedia.org/w/index.php?title=Role&oldid=307956739, 4 September 2009.

[INT 37] http://www.answers.com/topic/cybernetics, 4 September 2009.

[INT 38] http://www.answers.com/topic/feedback, 4 September 2009.

[INT 39] http://www.cultsock.ndirect.co.uk/MUHome/cshtml/introductory/sw.html. No Longer Available Online. Last Access: 23 March 2009.

[INT 40] http://en.wikipedia.org/w/index.php?title=Interactivity&oldid=275234399, 4 September 2009.

[INT 41] http://en.wikipedia.org/w/index.php?title=Compiler&oldid=278869644, 4 September 2009.

[INT 42] http://en.wikipedia.org/w/index.php?title=Batch_processing&oldid=279158034, 4 September 2009.

[INT 43] http://en.wikipedia.org/w/index.php?title=Theories_of_technology&oldid=271686124, 4 September 2009.

[INT 44] http://www.worldofwarcraft.com/, 4 September 2009.

[INT 45] http://secondlife.com/, 4 September 2009.

[INT 46] http://www.interaction-design.org/encyclopedia/context_of_use.html, 4 September 2009.

[INT 47] http://www.usability.gov/basics/whatusa.html, 4 September 2009.

[INT 48] http://www.cooper.com/newsletters/2002_05/dont_get_burned_by_bad_mapping.htm, 4 September 2009.

[INT 49] http://en.wikipedia.org/w/index.php?title=Metacognition&oldid=267341659, 4 September 2009.

[INT 50] http://www.imdb.com/character/ch0000574/, 4 September 2009.

[INT 51] http://www.imdb.com/character/ch0000630/, 4 September 2009.

[INT 52] http://www.unesco.org/education/imld_2002/unversal_decla.shtml, 4 September 2009.

[INT 53] http://support.sony-europe.com/aibo/, 4 September 2009.

[INT 54] http://en.wikipedia.org/w/index.php?title=Petz&oldid=310405684, 4 September 2009.

[INT 55] http://en.wikipedia.org/w/index.php?title=Furby&oldid=307623830, 4 September 2009.

[INT 56] http://en.wikipedia.org/w/index.php?title=Tamagotchi&oldid=310231573, 4 September 2009.

[INT 57] http://www.microsoft.com/, 4 September 2009.

[INT 58] http://research.microsoft.com/~horvitz/lum.htm, 4 September 2009.

[INT 59] http://www.imdb.com/title/tt0113568/, 4 September 2009.

[INT 60] http://www.disappearing-computer.net/, 4 September 2009.

[INT 61] http://www.anthropomorphism.org/, 4 September 2009.

[INT 62] http://www.tombraider.com/, 4 September 2009.

[INT 63] http://www.writerswrite.com/journal/jun98/lazy2.htm, 4 September 2009.

[INT 64] http://www.imdb.com/title/tt0114709/, 4 September 2009.

[INT 65] http://www.imdb.com/title/tt0120587/, 4 September 2009.

[INT 66] http://www.imdb.com/title/tt0126029/, 4 September 2009.

[INT 67] http://www.imdb.com/title/tt0107290/, 4 September 2009.

[INT 68] http://www.imdb.com/title/tt0120915/, 4 September 2009.

[INT 69] http://www.imdb.com/title/tt0121765/, 4 September 2009.

[INT 70] http://www.imdb.com/title/tt0121766/, 4 September 2009.

[INT 71] http://www.imdb.com/title/tt0120737/, 4 September 2009.

[INT 72] http://www.imdb.com/title/tt0167261/, 4 September 2009.

[INT 73] http://www.imdb.com/title/tt0167260/, 4 September 2009.

[INT 74] http://www.imdb.com/title/tt0173840/, 4 September 2009.

[INT 75] http://en.wikipedia.org/w/index.php?title=Final_Fantasy:_The_Spirits_Within&oldid=97317491, 4 September 2009.

[INT 76] http://us.aving.net/news/view.php?articleId=31719, 4 September 2009.

[INT 77] http://www.kois.go.kr/news/news/newsView.asp?serial_no=20061016001&part=107&SearchDay=&page=1. No Longer Available Online. Last Access: 24 November 2007.

[INT 78] https://listserv.temple.edu/cgi-bin/wa?A2=ind0610c&L=presence-l&O=D&P=1360, 4 September 2009.

[INT 79] http://www.web20asia.com/tag/robot, 6 October 2009.

[INT 80] http://asia.cnet.com/reviews/blog/seouledout/0,39061071,61961756,00.htm, 4 September 2009.

[INT 81] http://blog.wired.com/gadgets/2007/01/ever2_muse_the_.html, 4 September 2009.

[INT 82] http://www.kuroneko-chan.com/echoes/index.php?s=baeg, 4 September 2009.

[INT 83] http://news.bbc.co.uk/2/hi/technology/6376479.stm, 4 September 2009.

[INT 84] http://www.experiencefestival.com/a/Ethnic_group_-_Ethnicity/id/1273798, 4 September 2009.

[INT 85] http://www.statmt.org/, 4 September 2009.

[INT 86] http://www.smartkom.org/eng/project_en_frames.pl?media_en.html, 4 September 2009.

[INT 87] http://en.wikipedia.org/w/index.php?title=Modality_%28human-computer_interaction%29&oldid=272098880, 4 September 2009.

[INT 88] http://en.wikipedia.org/w/index.php?title=Video_tracking&oldid=273638671, 4 September 2009.

[INT 89] http://en.wikipedia.org/w/index.php?title=Eye_tracking&oldid=282714461, 4 September 2009.

[INT 90] http://www.linguistics-online.com/, Module Conversation – An Overview, 4 September 2009.

[INT 91] http://www.loebner.net/Prizef/loebner-prize.html, 4 September 2009.

[INT 92] http://en.wikipedia.org/w/index.php?title=Teleprinter&oldid=288999760, 4 September 2009.

[INT 93] http://en.wikipedia.org/w/index.php?title=MUD&oldid=289554459, 4 September 2009.

[INT 94] http://alicebot.blogspot.com/, 4 September 2009.

[INT 95] http://www.botspot.com/pages/chatbots.html, 4 September 2009.

[INT 96] http://www.chatterboxchallenge.com, 4 September 2009.

[INT 97] http://www.a-i.com/show_tree.asp?id=61&level=3&root=1, 4 September 2009.

[INT 98] http://en.wikipedia.org/w/index.php?title=Joseph_Weizenbaum&oldid=283972523, 4 September 2009.

[INT 99] http://www.geocities.com/rorysr2002/, 4 September 2009.

[INT 100] http://en.wikipedia.org/w/index.php?title=Mirroring_%28psychology%29&oldid=260909996, 4 September 2009.

[INT 101] http://en.wikipedia.org/w/index.php?title=Expert&oldid=293057511, 4 September 2009.

[INT 102] http://en.wikipedia.org/w/index.php?title=Knowledge&oldid=294659634, 4 September 2009.

[INT 103] http://en.wikipedia.org/w/index.php?title=Epistemology&oldid=292536300, 4 September 2009.

[INT 104] http://en.wikipedia.org/w/index.php?title=Belief&oldid=294348705, 4 September 2009.

[INT 105] http://en.wikipedia.org/w/index.php?title=Descriptive_knowledge&oldid=283589076, 4 September 2009.

[INT 106] http://en.wikipedia.org/w/index.php?title=Procedural_knowledge&oldid=283589939,

4 September 2009.

[INT 107] http://en.wikipedia.org/w/index.php?title=Domain_knowledge&oldid=206995703, 4 September 2009.

[INT 108] http://en.wikipedia.org/w/index.php?title=Commonsense_knowledge_base &oldid=261229821, 4 September 2009.

[INT 109] http://en.wikipedia.org/w/index.php?title=Explicit_knowledge&oldid=291398590, 4 September 2009.

[INT 110] http://en.wikipedia.org/w/index.php?title=Tacit_knowledge&oldid=293718892, 4 September 2009.

[INT 111] http://www.epistemics.co.uk/Notes/63-0-0.htm, 4 September 2009.

[INT 112] http://en.wikipedia.org/w/index.php?title=Knowledge_base&oldid=288736039, 4 September 2009.

[INT 113] http://en.wikipedia.org/w/index.php?title=Expert_system&oldid=293031758, 4 September 2009.

[INT 114] http://www.epistemics.co.uk/Notes/174-0-0.htm, 4 September 2009.

[INT 115] http://www.epistemics.co.uk/Notes/90-0-0.htm, 4 September 2009.

[INT 116] http://www.epistemics.co.uk/Notes/178-0-0.htm, 4 September 2009.

[INT 117] http://www.epistemics.co.uk/Notes/188-0-0.htm, 4 September 2009.

[INT 118] http://www.epistemics.co.uk/Notes/186-0-0.htm, 4 September 2009.

[INT 119] http://www.epistemics.co.uk/Notes/184-0-0.htm, 4 September 2009.

[INT 120] http://www.epistemics.co.uk/Notes/180-0-0.htm, 4 September 2009.

[INT 121] http://en.wikipedia.org/w/index.php?title=Cluster_analysis&oldid=295204195, 4 September 2009.

[INT 122] http://en.wikipedia.org/w/index.php?title=Genetic_algorithm&oldid=295828902, 4 September 2009.

[INT 123] http://en.wikipedia.org/w/index.php?title=Fuzzy_logic&oldid=295913934, 4 September 2009.

[INT 124] http://en.wikipedia.org/w/index.php?title=Situation_calculus&oldid=263338130, 4 September 2009.

[INT 125] http://en.wikipedia.org/w/index.php?title=Object-oriented_programming &oldid=298744497, 4 September 2009.

[INT 126] http://www.w3.org/2001/tag/doc/selfDescribingDocuments.html, 4 September 2009.

[INT 127] http://www.w3.org/2001/sw/, 4 September 2009.

[INT 128] http://www.w3.org/XML/, 4 September 2009.

[INT 129] http://www.w3.org/RDF/, 4 September 2009.

[INT 130] http://www.w3.org/2004/OWL/, 4 September 2009.

[INT 131] http://www.w3.org/TR/rdf-schema/, 4 September 2009.

[INT 132] http://www.ruleml.org/, 4 September 2009.

[INT 133] http://www.w3.org/Submission/SWRL/, 4 September 2009.

[INT 134] http://www.w3.org/2002/Talks/04-sweb/slide12-0.html, 4 September 2009.

[INT 135] http://www.w3.org/2004/Talks/0412-RDF-functions/slide4-0.html, 4 September 2009.

[INT 136] http://vanderwal.net/folksonomy.html, 4 September 2009.

[INT 137] http://en.wikipedia.org/w/index.php?title=Digital_signature&oldid=303904964, 4 September 2009.

[INT 138] http://en.wikipedia.org/w/index.php?title=Semantic_Web&oldid=304602838, 4 September 2009.

[INT 139] http://en.wikipedia.org/w/index.php?title=Means-ends_analysis&oldid=297867161, 4 September 2009.

[INT 140] http://en.wikipedia.org/w/index.php?title=Reactive_planning&oldid=281609025, 4 September 2009.

[INT 141] http://en.wikipedia.org/w/index.php?title=Naive_Bayes_classifier&oldid=304676111, 4 September 2009.

[INT 142] http://en.wikipedia.org/w/index.php?title=Conditional_probability&oldid=299352946, 4 September 2009.

[INT 143] http://en.wikipedia.org/w/index.php?title=Prior_probability&oldid=298238979, 4 September 2009.

[INT 144] http://en.wikipedia.org/w/index.php?title=Collaborative_filtering&oldid=300591289, 4 September 2009.

[INT 145] http://en.wikipedia.org/w/index.php?title=Bayesian_network&oldid=136159891, 4 September 2009.

[INT 146] http://www.cs.mdx.ac.uk/staffpages/serengul/Tutorial/Overlay.student.models.htm, 4 September 2009.

[INT 147] http://www.cs.mdx.ac.uk/staffpages/serengul/Tutorial/Differential.student.models.htm, 4 September 2009.

[INT 148] http://www.cs.mdx.ac.uk/staffpages/serengul/Tutorial/Perturbation.student.models.htm, 4 September 2009.

[INT 149] http://www.encyclo.co.uk/define/scaffold, 4 September 2009.

[INT 150] http://en.wikipedia.org/w/index.php?title=Test_%28student_assessment%29&oldid=309748585, 4 September 2009.

[INT 151] http://en.wikipedia.org/w/index.php?title=Assessment&oldid=306155338, 4 September 2009.

[INT 152] http://en.wikipedia.org/w/index.php?title=Motivation&oldid=310270851, 4 September 2009.

[INT 153] http://en.wikipedia.org/w/index.php?title=Peer_tutor&oldid=266388429, 4 September 2009.

[INT 154] http://www.imdb.com/title/tt0063951/, 4 September 2009.

[INT 155] http://en.wikipedia.org/w/index.php?title=William_James&oldid=268163037, 4 September 2009.

[INT 156] http://en.wikipedia.org/w/index.php?title=Flow_%28psychology%29&oldid=124187558, 4 September 2009.

[INT 157] http://en.wikipedia.org/w/index.php?title=Facial_feedback_hypothesis&oldid=261907976, 4 September 2009.

[INT 158] http://face-and-emotion.com/dataface/facs/description.jsp, 4 September 2009.

[INT 159] http://en.wikipedia.org/w/index.php?title=Automated_theorem_proving&oldid=304514608, 4 September 2009.

[INT 160] http://en.wikipedia.org/w/index.php?title=Blackboard_system&oldid=173532606, 4 September 2009.

[INT 161] http://www.adobe.com/, 4 September 2009.

[INT 162] http://www.microsoft.com/products/msagent/main.aspx, 4 September 2009.

[INT 163] http://www.cantoche.com/, 4 September 2009.

[INT 164] http://www.microsoft.com/windows/, 4 September 2009.

[INT 165] http://www.apple.com/macosx/, 4 September 2009.

[INT 166] http://www.kernel.org/, 4 September 2009.

[INT 167] http://www.w3.org/html/, 4 September 2009.

[INT 168] https://developer.mozilla.org/en/JavaScript, 4 September 2009.

[INT 169] http://java.sun.com/, 4 September 2009.

[INT 170] http://www.dcs.gla.ac.uk/~johnson/teaching/isd/course.html, 4 September 2009.

[INT 171] http://www.nationmaster.com/encyclopedia/Requirements-analysis, 4 September 2009.

[INT 172] http://en.wikipedia.org/w/index.php?title=Software_prototyping&oldid=268637341, 4 September 2009.

[INT 173] http://en.wikipedia.org/w/index.php?title=Wizard_of_Oz_experiment&oldid=262250803, 4 September 2009.

[INT 174] http://en.wikipedia.org/w/index.php?title=Waterfall_model&oldid=271759527, 4 September 2009.

[INT 175] http://en.wikipedia.org/w/index.php?title=Iterative_and_incremental_development &oldid=268788464, 4 September 2009.

[INT 176] http://www.techweb.com/encyclopedia/defineterm.jhtml?term=iterative+development, 4 September 2009.

[INT 177] http://manta.cs.vt.edu/cs3704/SEmodule/SE/Lessons/Spiral/index.html, 4 September 2009.

[INT 178] http://www.useit.com/papers/heuristic/heuristic_list.html, 4 September 2009.

[INT 179] http://en.wikipedia.org/w/index.php?title=Cognitive_walkthrough&oldid=253908968, 4 September 2009.

[INT 180] http://www.pages.drexel.edu/~zwz22/CognWalk.htm, 4 September 2009.

[INT 181] http://www.imdb.com/title/tt0032138/, 4 September 2009.

[INT 182] http://wwwhome.cs.utwente.nl/~zsofi/eeca/, 4 September 2009.

[INT 183] http://en.wikipedia.org/w/index.php?title=Falsifiability&oldid=182026152, 4 September 2009.

[INT 184] http://en.wikipedia.org/w/index.php?title=Hypothesis&oldid=185234775, 4 September 2009.

[INT 185] http://www.theresearchassistant.com/tutorial/2-1.asp, 4 September 2009.

[INT 186] http://en.wikipedia.org/w/index.php?title=Operationalization&oldid=271370444, 4 September 2009.

[INT 187] http://en.wikipedia.org/w/index.php?title=Construct_validity&oldid=181707720, 4 September 2009.

[INT 188] http://en.wikipedia.org/w/index.php?title=Scientific_method&oldid=273395490, 4 September 2009.

[INT 189] http://en.wikipedia.org/w/index.php?title=Statistical_survey&oldid=184227452, 4 September 2009.

[INT 190] http://www.holah.karoo.net/experimental_method.htm, 4 September 2009.

[INT 191] http://en.wikipedia.org/w/index.php?title=Experiment&oldid=273383023, 4 September 2009.

[INT 192] http://en.wikipedia.org/w/index.php?title=Case_study&oldid=157004583, 4 September 2009.

[INT 193] http://www.dobney.com/Intelligence/desk_research.htm, 4 September 2009.

[INT 194] http://en.wikipedia.org/w/index.php?title=Level_of_measurement&oldid=273360571, 4 September 2009.

[INT 195] http://www.uoguelph.ca/htm/MJResearch/ResearchProcess/741process10a2.htm, 4 September 2009.

[INT 196] http://www.uoguelph.ca/htm/MJResearch/ResearchProcess/IntervalScale.htm, 4 September 2009.

[INT 197] http://www.uoguelph.ca/htm/MJResearch/ResearchProcess/RatioScale.htm, 4 September 2009.

[INT 198] http://www.epa.gov/evaluate/glossary/s-esd.htm, 4 September 2009.

[INT 199] http://www.ucd.ie/teaching/teachingEvaluation_sub/student_evaluations.html, 4 September 2009.

[INT 200] http://en.wikipedia.org/w/index.php?title=Participant_observation&oldid=271384398, 4 September 2009.

[INT 201] http://en.wikipedia.org/w/index.php?title=Data_analysis&oldid=182632723, 4 September 2009.

[INT 202] http://www.wilderdom.com/research/QualitativeVersusQuantitativeResearch.html, 4 September 2009.

[INT 203] http://en.wikipedia.org/w/index.php?title=Descriptive_statistics&oldid=272270951, 4 September 2009.

[INT 204] http://hsc.uwe.ac.uk/dataanalysis/quantDescCentral.asp, 4 September 2009.

[INT 205] http://hsc.uwe.ac.uk/dataanalysis/quantDescDisp.asp, 4 September 2009.

[INT 206] http://en.wikipedia.org/w/index.php?title=Standard_deviation&oldid=273390791, 4 September 2009.

[INT 207] http://en.wikipedia.org/w/index.php?title=Statistical_dispersion&oldid=146869903, 4 September 2009.

[INT 208] http://en.wikipedia.org/w/index.php?title=Statistical_inference&oldid=272282151, 4 September 2009.

[INT 209] http://en.wikipedia.org/w/index.php?title=Statistics&oldid=272803011, 4 September 2009.

[INT 210] http://www.stats.gla.ac.uk/steps/glossary/confidence_intervals.html, 4 September 2009.

[INT 211] http://hsc.uwe.ac.uk/dataanalysis/quantIssuesHyp.asp, 4 September 2009.

[INT 212] http://en.wikipedia.org/w/index.php?title=Alternate_hypothesis&oldid=175785606, 4 September 2009.

[INT 213] http://en.wikipedia.org/w/index.php?title=Novelty_effect&oldid=205063885, 4 September 2009.

[INT 214] http://www.chip.de/news/Microsoft-toetet-Weihnachtsmann-nach-Sex-Affaere_29740137.html, 4 September 2009.

[INT 215] http://www.ims.uni-stuttgart.de/~moehler/synthspeech/, 4 September 2009.