

No. 15-2011

Jürgen-Peter Kretschmer

**The Sequencing Problem in Sequential Investigation
Processes**

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

The Sequencing Problem in Sequential Investigation Processes

Jürgen-Peter Kretschmer*

Abstract: Many decision problems in various fields of application can be characterized as diagnostic problems trying to assess the true state (of the world) of given cases. The investigation of assessment criteria improves the initial information according to observed signal outcomes, which are related to the possible states. Such sequential investigation processes can be analyzed within the framework of statistical decision theory, in which prior probability distributions of classes of cases are updated, allowing for a sorting of particular cases into ever smaller subclasses. However, receiving such information causes investigation costs. Besides the question about the set of relevant criteria, this defines two additional problems of statistical decision problems: the optimal stopping of investigations and the optimal sequence of investigating a given set of criteria. Unfortunately, no solution exists with which the optimal sequence can generally be determined. Therefore, the paper characterizes the associated problems and analyzes existing heuristics trying to approximate an optimal solution.

JEL-Classification: D80, D81, C11, C44

Keywords: Decision-Making, Uncertainty, Information, Bayesian Analysis, Statistical Decision Theory

1. Introduction

Many decision problems can be characterized as diagnostic problems, in which decisionmakers try to determine the correct consequences for cases they have to decide on. The determination generally follows an assessment process of the true state of the world, by which the diagnosis is improved pursuant to realizations of the investigation of specific features.

For example, physicians are interested in the presence or absence of certain diseases to match medical treatment. Performing a number of medical tests enables them to refine their initial diagnosis and respective intended treatment, according to the observed outcomes of the

* Philipps-University Marburg, Department of Business Administration and Economics, Economic Policy Unit, Am Plan 2, D-35032 Marburg, Germany, Phone: +49-6421-2823925, Fax: +49-6421-2823936, E-mail: kretschmer@wiwi.uni-marburg.de

I thank Georg von Wangenheim for helpful comments.

specific tests. Competition authorities are interested in whether certain business behaviors (e.g. specific types of agreements or horizontal mergers) are characterized by a positive or negative welfare effect (i.e. whether they are pro- or anticompetitive). The diagnosis is refined by the assessment of different market indicators. Subsequently, decisions of permission or prohibition follow from specific outcomes of these indicators. In educational questions – such as examinations – as well as in psychological issues one is concerned about the correct assessment of specific levels of knowledge. Conducting certain tests allows for a refinement of the diagnosed level of knowledge, which is based on the observed answers.

All such decision situations have in common both that it is *ex ante* uncertain, which of the possible states of the world is realized in concrete cases, and an assessment process aiming at the improvement of the given imperfect information. The stepwise refinement of an initial diagnosis by sequentially investigating certain features and observing specific outcomes can be represented by a multi-stage decision tree. At each stage, decisionmakers have to determine the next relevant feature and whether the benefits of further investigation outweigh additional costs. Thus, termination at early stages may become preferable because further information does not alter the terminal decision and/or is too costly to gather.

Such kinds of problems are statistical decision problems,¹ referring to the classification of different single decision situations according to recurring or corresponding characteristics.² Accordingly, the *ex ante* uncertainty about a class of cases is expressed by a (subjective) probability distribution over the possible states of the world – based on experience and theoretical analyses. This information can be improved by investigation of different criteria (or tests) out of a given set, the outcomes of which somehow relate to the possible states of the world. For example, reacting to certain medical tests in a specified way (or showing specific symptoms) increases the probability of a certain disease (Gorry 1967: 23; Warner/Veasey 1992: 43 et seqq.), observing low market shares reduces the potential of exerting market power and the probability of anti-competitive vertical restraints (Motta 2004: 117-118, 377-378), and answering to certain tests in a correct way increases the probability of being a high-knowledge student (Vos 1999: 273; 2007: 617). Thus, the investigation of criteria leads to a refinement of the distributions and a sorting of particular cases into ever smaller subclasses (Gorry 1967: 23-24), with any given history of outcomes assigning specific probabilities to the possible states of the world. Accordingly, general diagnostic problems (which can be found in lots of different areas) can be described by so-called differentiated multi-stage decision processes with different possible investigation paths. The latter are characterized by different developments of probability distributions.

The decision in favor of one of the possible states of the world over another is accompanied with some probability of decision failure and its possible consequence (error costs). In medical diagnosis problems, the possible consequences are represented by the costs of misdiagnosis (Gorry/Barnett 1968: 494), which comprise the misdiagnosis of one disease as another or the failing in identifying a specific disease. The costs of misdiagnosis most notably

¹ See, for example, DeGroot (1970: 136 et seqq.) for the general definition of the essential components of a statistical decision problem.

² See Vanberg (1993: 175) for a more detailed description of decision situation regularities.

include the consequences associated with inappropriate uses of health services, avoidable or failed treatments as well as illness and distress related (emotional) factors (Reid et al. 1998: 354). When regarding competition policy concerns, losses from permitting or prohibiting certain business behaviors may occur – among other things – as supra-competitive pricing, prevention of innovative activities or abuse of market power. And in educational questions, losses from terminal decisions on the level of knowledge include all relevant psychological, social and economic consequences (Vos 2007: 613). Investigating different criteria aims at reducing the uncertainty about the true state of the world and the resulting error costs. Since the investigation of any criterion raises costs (investigation costs) the improvement of information (i.e. reduction of error costs) might not be justified. The comparison of expected total costs (i.e. the sum of error and investigation costs) of an immediate decision with those obtained if more information is gathered might terminate the assessment process prior to the investigation of every possible criterion. This problem of optimal stopping will only be addressed shortly in this paper.

Besides, there is the effect of re-arranging the sequence of criteria, on which the paper focuses: Even though total investigation costs remain unaffected, changing the sequence of criteria has a side-effect on the potential of reducing error costs of every (other) criterion (Mussi 2002: 103-104). Stated differently, information gathered by the investigation of criteria differs according to the reference point of current knowledge. Therefore, the actually realized criteria on every investigation path depend (or may depend) on the investigated sequence. For example, the significance of certain reactions to a specific medical test depends on previously conducted tests and observed reactions; certain examination questions provide for a ‘fine-tuning’ of knowledge levels at later stages, but fail to provide meaningful information when applied in examinations early. The optimal sequence simply minimizes total expected costs, i.e. aggregated error and investigation costs. Unfortunately, no solution exists with which the optimal sequence of criteria can generally be determined; therefore we have to refer to certain heuristics. The analysis does not refer to special fields of interest, but rather allows for a more general application.

This paper is the first comprehensive overview of the sequencing problem in sequential investigation processes. Present literature comprises a variety of articles about different aspects, but lacks a systematic analysis. Next to summarizing the determinants of the sequencing problem and surveying so far offered heuristic solutions, this paper further enhances the number of solutions offered by combining the myopic heuristic with a dominance characteristic of certain criteria. The paper is structured as follows: Section 2 contains a simple model of sequential investigation processes. Heuristic solutions are analyzed and discussed in section 3. Addressing specific constellations among the characteristics of investigation processes allows for tentative conclusions concerning the advantageousness of one heuristic or another. Finally, conclusions are drawn in section 4.

2. Optimizing the sequence of assessment criteria in sequential investigation processes

Sequential investigation processes characterize general diagnostic problems, which try to determine the right consequences for specific cases by improving the ex ante uncertainty about their true state of the world. This is accomplished by the sequential assessment of criteria allowing for a refinement of underlying expectations about the frequencies of different possible states of the world if certain characteristics are observed after their respective assessment. The first problem arises in determining the relevant set of criteria. In general, the existence of an exhausting amount of possible criteria may be far from obvious; furthermore, there may be disagreements about the ability of certain criteria in improving the respective decision basis. Even though these problems arise in the first place and should be dealt with before the analysis of a sequential process starts, it is not subject of this paper. Assuming that the relevant set of criteria is defined, the following analysis introduces the second problem of sequential investigation processes, which is the optimal amount of gathered information. The next section offers the basic model and determines when final decisions should optimally be handed down without further information. This determination is needed to analyze the third problem of sequential investigation processes, which is the main subject of this paper, i.e. determining the optimal sequence in which a set of criteria optimally should be investigated. Answering the question of determining optimal sequential investigation processes requires solving these three central problems.

2.1 Basic Model

The analysis of the sequencing problem within sequential investigation processes is conducted by illustrating and using the model of Kerber et al. (2008: 5-11), who mainly referred to competition law as the field of application.³

Starting point is a decisionmaker trying to assess different cases out of a given class of cases concerning their true states of the world.⁴ The decisionmaker has some knowledge – based on theoretical analyses and experience – about the prior probability distribution of the possible states of the world w and their respective (quantitative) consequences. Assuming that the possible states of the world are dichotomous, $w \in \{B, H\}$, the prior probabilities (priors) of $w = B$ ('beneficial') and $w = H$ ('harmful') are given by p and $1 - p$, respectively.⁵

³ The following sequential decision model simplifies general designs of statistical decision theory to a certain extent, without loss of generality (see, e.g., Berger (1993), DeGroot (1970), Hadley (1967), Hamburg (1970) or Moore/Whinston (1986 and 1987) for more general analyses).

⁴ Such a classification refers to corresponding characteristics of different single decision situations (Vanberg 1993: 175). For example, a class of cases may be given by resale price maintenance or horizontal mergers in competition law, students of the same semester and/or lecture when assessing a certain knowledge-level, or patients in whom congenital heart disease is suspected.

⁵ The analysis can also be conducted for a range of welfare effects with a continuous prior probability density function. However, besides complicating the analysis, the results and implications remain the same.

This *a-priori* information can be refined by the investigation of k distinct assessment criteria, labeled by the index $i \in K = \{1, 2, \dots, k\}$, whose outcomes are related to the two states of the world and function as evidence for the more probable occurrence of one of these states. The investigation of those criteria is assumed to be without error and does not alter the set of possible states of the world. For each of these criteria, the decisionmaker may gather exactly one signal z_i , whose outcome $m_i \in M \equiv \{-1, +1\}$ is dichotomous.⁶ The conditional probabilities (likelihoods) of observing $m_i = +1$ and $m_i = -1$ are given by q_i^w and $1 - q_i^w$, respectively, for the true state of the world w .⁷

When assuming that the decisionmaker may collect only one signal at a time the investigation process is split into stages $j \in \{0, 1, \dots, k\}$, where j denotes the number of signals already collected. In fact, the likelihoods of the different criteria may depend on already collected signals and their outcomes – otherwise the signals are stochastically independent. Thus, after having collected j signals, the decisionmaker knows a history h_j of collected signals and their outcomes. This history consists of a sequence of j pairs $(i_j, m_j) \in K \times \{-1, +1\}$, in which the first entry denotes the index of the signal collected at stage \tilde{j} and the second entry denotes the outcome of this signal. For any given history h_j the decisionmaker may then use the conditional likelihoods $q_i^w(h_j)$ to determine updated probabilities (posteriors) of the true state of the world, denoted by $\hat{p}(h_{j+1})$ and $1 - \hat{p}(h_{j+1})$, by iterated Bayesian updating. Assuming that signal z_i is collected at stage j and writing h_j^m for a history h_j whose last signal yielded outcome $m \in \{-1, +1\}$, Bayesian updating yields

$$(1) \quad \hat{p}(h_{j+1}^{+1}) = \frac{\hat{p}(h_j) q_i^B(h_j)}{q_i(h_j)} \quad \text{and} \quad \hat{p}(h_{j+1}^{-1}) = \frac{\hat{p}(h_j) (1 - q_i^B(h_j))}{(1 - q_i(h_j))},$$

with

$$(2) \quad q_i(h_j) = \hat{p}(h_j) q_i^B(h_j) + (1 - \hat{p}(h_j)) q_i^H(h_j) \quad \text{and} \\ 1 - q_i(h_j) = \hat{p}(h_j) (1 - q_i^B(h_j)) + (1 - \hat{p}(h_j)) (1 - q_i^H(h_j))$$

as the probabilities of observing the outcomes $m_i = +1$ and $m_i = -1$, respectively, when the true state of the world is unknown (unconditional probabilities). Generally, the unconditional

⁶ The interpretation of the signal outcomes can differ: for example, specific symptoms may be absent or not, the answer to an examination question may be ‘right’ or ‘wrong’, market shares may be ‘high’ or ‘low’, and a business practice may be initiated by producers or retailers.

⁷ In general, these likelihoods may differ between different decisionmakers, thereby reflecting their ability to discriminate between the possible states of the world by investigation of the respective criteria. Depending on the concrete decision situation, the likelihoods at hand can be recalculated after a sufficient number of observations of the realized states of the world (Warner/Veasey 1992: 50). The improvement of the underlying criteria, respectively ‘evidence system’, can then be used for future decision situations.

probability of observing the outcome m_i at stage j is the sum of the conditional probabilities of observing the outcome m_i , weighted by the posteriors of stage j (Mag 1977: 234-235). The definition of $\hat{p}(h_0) \equiv p$ and $q_i^w(h_0)$ as the likelihoods without any history of observed signals terminates the iterative updating.

To avoid trivial results, the signals are assumed to be neither totally uninformative nor perfectly informative. The former condition requires the more probable inference – at least to some extent – to one of the two states of the world when observing a specific outcome, i.e. the likelihood that the outcome of signal z_i is m_i is not the same for both states of the world ($q_i^H \neq q_i^B$). As a result (see (1) and (2)), the updated probabilities at stage $j+1$ differ from those at stage j ($\hat{p}(h_{j+1}^{m_i}) \neq \hat{p}(h_j)$). Assuming that $q_i^B(h_j) > q_i^H(h_j)$, i.e. the outcome $m_i = +1$ ($m_i = -1$) is more likely for the beneficial (harmful) state of the world, the posteriors then fulfill: $\hat{p}(h_{j+1}^{+1}) > \hat{p}(h_j) > \hat{p}(h_{j+1}^{-1})$. The latter condition excludes the perfect inference on one of the two states of the world when observing a specific outcome ($0 < \hat{p}(h_{j+1}^{m_i}) < 1$). Both conditions can thus be summarized as follows: $0 < q_i^H \neq q_i^B < 1$. These likelihoods can be used to characterize the different criteria with respect to their power in discriminating between the two states of the world – as revealed, e.g., by the ratio q_i^B/q_i^H (Heckerman et al. 1989: 6).

Costs accrue in two ways: first, from collecting more information (investigation costs), and second, from making wrong decisions (error costs). Investigation costs of collecting signal z_i are independent of all other criteria and given by $c_i > 0$. For ease of exposition, we will assume that $c_i = c \forall i \in K$, unless otherwise stated. Error costs are determined by a loss function $L(w, d)$, which represents the loss when the true state of the world is w and the decisionmaker chooses decision $d \in \{d^+, d^-\}$, where d^+ and d^- are two terminal decision possibilities (e.g., permission and prohibition or passing and failing) with d^+ (d^-) relating to $w = B$ ($w = H$). Normalizing the loss from making correct decisions to zero ($L(B, d^+) = L(H, d^-) = 0$), we get $L(H, d^+) = L_H$ and $L(B, d^-) = L_B$. Total expected costs (EC) from terminating the investigation after history h_j are thus given by:

$$(3) \quad EC(h_j, d^+) = jc + EL(h_j, d^+) = jc + L_H(1 - \hat{p}(h_j)) \text{ and}$$

$$EC(h_j, d^-) = jc + EL(h_j, d^-) = jc + L_B \hat{p}(h_j),$$

with EL as the expected loss associated with the respective decision.⁸ This allows representing the sequential investigation structure by a multi-stage decision tree (see Kerber et al. 2008: 6).

⁸ In the case of a range of possible states of the world, total expected costs are given by:

When first analyzing the optimal depth of investigation we will take the sequence of criteria within the decision tree as given. For simplicity, we assume that signal i is collected at stage j . At each stage except for the last, the decisionmaker has to decide from a trichotomous choice set $D_j = \{d^+, d^-, d^0\}$, which comprises two terminating decisions (d^+ and d^-) and one continuing decision (collecting a $j+1^{\text{st}}$ signal, d^0). For the last stage the choice set comprises only the two terminating decisions. The decision is made by calculating the minimum of the expected total costs of each decision possibility.

Expected total costs for continuing the investigation after history h_j ($EC(h_j, d^0)$) depend on the signals and choices of all possible later stages (Lewis/Sheehan 1990: 374). We define $EC^*(h_j) = \min_{D_j} [EC(h_j, d^+), EC(h_j, d^-), EC(h_j, d^0)]$ as the minimal expected total costs after history h_j , and $d^*(h_j)$ as the optimal decision achieving $EC^*(h_j)$. Hence, $EC(h_j, d^0)$ is given by the expected value of the minimal expected total costs of the next stage, taken over the unconditional probabilities of the next signal (Marschak/Miyasawa 1968: 144):

$$q_i(h_j) \text{ for } m_i = +1 \text{ and } 1 - q_i(h_j) \text{ for } m_i = -1.$$

Thus,

$$(4) \quad EC(h_j, d^0) = q_i(h_j) EC^*(h_{j+1}^+) + (1 - q_i(h_j)) EC^*(h_{j+1}^-).$$

Specifying criteria by means of the likelihoods and Bayes' theorem (see (1)) or directly by means of the posterior probabilities and their possible unconditional probabilities is equivalent (Ponssard 1975: 449). We thus can reverse the formulae of Bayesian updating given in (1) to express the likelihoods and unconditional probabilities solely in terms of prior and posterior probabilities of the possible states of the world.¹⁰ Given the consistency condition of the priors being equal to the average of the posteriors of the next stage (Athey/Levin 2000: 3), $q_i(h_j) \hat{p}(h_{j+1}^+) + (1 - q_i(h_j)) \hat{p}(h_{j+1}^-) = \hat{p}(h_j)$, the unconditional probabilities can be expressed as follows:

$$(5) \quad q_i(h_j) = \frac{\hat{p}(h_j) - \hat{p}(h_{j+1}^-)}{\hat{p}(h_{j+1}^+) - \hat{p}(h_{j+1}^-)} \quad \text{and} \quad 1 - q_i(h_j) = \frac{\hat{p}(h_{j+1}^+) - \hat{p}(h_j)}{\hat{p}(h_{j+1}^+) - \hat{p}(h_{j+1}^-)}.$$

The reversion, however, only holds for the general discrete case if the number of possible states of the world at least equals the number of possible signal outcomes (see Appendix A).

$$EC(h_j, d^+) = jc + \int_{w < 0} L(w, d^+) \hat{p}(h_j) dw \quad \text{and} \quad EC(h_j, d^-) = jc + \int_{w > 0} L(w, d^-) \hat{p}(h_j) dw,$$

with $L(w, d)$ as a continuous loss function and $\hat{p}(h_j)$ as the probability density function of history h_j .

⁹ It is assumed for simplicity that (quantitative) losses are the same for each possible stage, whereas in practice the loss structure may alter due to delays in reaching terminating decisions (Hamburg 1970: 727).

¹⁰ Reversing the formulae of Bayesian updating in (1) at first gives the likelihoods in the following terms:

$$q_i^s(h_j) = \frac{\hat{p}(h_{j+1}^+)}{\hat{p}(h_j)} \Phi \quad \text{and} \quad q_i^u(h_j) = \frac{1 - \hat{p}(h_{j+1}^+)}{1 - \hat{p}(h_j)} \Phi \quad \text{with} \quad \Phi = \frac{\hat{p}(h_j) - \hat{p}(h_{j+1}^-)}{\hat{p}(h_{j+1}^+) - \hat{p}(h_{j+1}^-)}.$$

Optimal decisions and the consequential minimal expected total costs must then be determined in a backwards iterated manner, i.e. starting at stage k (Hamburg 1970: 733). At the final stage k , the optimal decision is given by

$$(6) \quad d_k^* = \arg \min_{d \in D_k} [EC(h_k, d)] \quad \text{where} \quad D_k = \{d^+, d^-\},$$

which is equivalent to selecting d^+ if and only if $\hat{p}(h_k) \geq L_H / (L_H + L_B)$.¹¹ Minimal expected total costs at stage k are thus given by

$$(7) \quad EC^*(h_k) = \min [k c + L_H (1 - \hat{p}(h_k)), k c + L_B \hat{p}(h_k)].^{12}$$

We can iterate backwards for $j < k$ by following

$$(8) \quad d_j^* = \arg \min_{d \in D_j} [EC(h_j, d)] \quad \text{where} \quad D_j = \{d^+, d^-, d^0\}$$

as decision rule. This yields

$$(9) \quad EC^*(h_j) = \min [j c + L_H (1 - \hat{p}(h_j)), j c + L_B \hat{p}(h_j), q_i(h_j) EC^*(h_{j+1}^+) + (1 - q_i(h_j)) EC^*(h_{j+1}^-)]$$

as minimal expected total costs at stage $j < k$.

This analysis determines how deep the investigation should continue in each of the branches of the multi-stage decision tree (optimal stopping). The decision rule given by (6) and (8) assigns to each observed outcome the optimal decision to be taken by the decisionmaker: Either terminating the investigation by deciding between d^+ and d^- or investigating another criterion. For a terminating decision being optimal in any of the branches, the reduction of expected losses (expected information value) achieved by additionally gathered information is not sufficient to justify the raised investigation costs, i.e. expected total costs are at their minimum. At any stage, this condition encompasses the entire subsequent decision tree and cannot be seen separately for only one (or the next) criterion, because the expected information value of any criterion is also affected by subsequent criteria and sometimes the reduction in expected losses of a subsequent criterion may be so high to over-compensate low expected information values of previous criteria (and the raised investigation costs). Appendix C addresses this characteristic because neglecting it is the crucial shortcoming of the so-called myopic approach.

¹¹ This decision rule assumes that the decisionmaker decides in favor of d^+ if indifferent between d^+ and other decisions. Only in the case of symmetric losses, the decisionmaker should decide in favor of the more probable state of the world. If, for example, L_H is assumed to be four times as large as L_B , the probability threshold equals $\hat{p}^B = 1/5$. The assumption of symmetric losses indicates equal undesirability of both types of possible erroneous decisions (Lewis/Sheehan 1990: 370; Beckner/Salop 1999: 61-62).

¹² If one of the terminating decisions is taken prior to stage k , one may replace k by any $j < k$ in (6) and (7).

2.2 Information value analysis

Given the definition of the minimum expected losses ($EL^*(h_j)$) analogously to $EC^*(h_j)$, the expected information value (IV) is defined by

$$(10) \quad IV_i(h_j) = EL^*(h_j) - \left[q_i(h_j) EL^*(h_{j+1}^{+1}) + (1 - q_i(h_j)) EL^*(h_{j+1}^{-1}) \right].$$

The expected information value is always non-negative (DeGroot 1962: 405-406; Gould 1974: 67; Heckerman et al. 1992: 97; Hirshleifer/Riley 1979: 1395) and can only be zero if the decisionmaker takes the same decision for both outcomes as taken without investigating the signal (Heckerman et al. 1991: 6; Hirshleifer/Riley 1992: 180). Taking the same decision with or without additional information cannot impair the decisionmaker's position in terms of expected losses (Cyert/DeGroot 1987: 16). This is formally ensured by the consistency condition of prior and updated probabilities (Athey/Levin 2001: 3; Hirshleifer/Riley 1992: 188), see (1) and (2): $\hat{p}(h_j) = q_i(h_j) \hat{p}(h_{j+1}^{+1}) + (1 - q_i(h_j)) \hat{p}(h_{j+1}^{-1})$.

A negative expected information value is generally excluded if additional information only affects the probabilities of the states of the world (Drukarczyk 1974: 9). It would only be possible if the posteriors of one state of the world change into the same 'direction' (increase or decrease) with respect to both possible signal outcomes; this case, however, is excluded directly by Bayes' theorem (see (1)).¹³

The non-negative expected information value must be compared to the strictly positive investigation costs in order to determine the expected total costs to be saved by investigating a criterion. In combination, both elements define the so-called *expected net information value*, which measures the informativeness of criteria in terms of reduced expected losses due to improved decision making and additional investigation costs (Kalagnanam/Henrion 1990: 272):¹⁴

$$(11) \quad NIV_i(h_j) = EC^*(h_j) - \left[q_i(h_j) EC^*(h_{j+1}^{+1}) + (1 - q_i(h_j)) EC^*(h_{j+1}^{-1}) \right].$$

Comparative statics of optimal stopping and realized investigation rules concerning the size of losses and investigation costs reveal the following: A general increase in investigation costs – while losses remain unchanged – tends to shorten the investigation process of any class; analogously, increasing the losses – while investigation costs remain unchanged – tends to deepen the investigation of any class. Increased losses increase the possible extent of the expected information value and thus call for a more accurate decision based on (an)

¹³ If additional information reduces the choice set (Drèze 1960) or becomes loss-relevant (i.e. changes the amount of losses), the expected information value may indeed become negative. The latter case of so-called 'signal dependence' is analyzed, e.g., by Bertocchi/Spagat (1998), Datta/ Mirman/Schlee (2000, 2002) and Sulganik/Zilcha (1997). However, these cases are not considered here. In the case of partial uncertainty about the required information inputs (states of the world, prior probabilities, choice set, loss function, signal outcomes, and likelihoods), Firschau (1980) analyzes bounds for the expected information value.

¹⁴ This definition of the information value as the difference of expected average values takes into account the quantitative effects (i.e. losses) as well (Howard 1966: 22). Other definitions, for example solely regarding the effects on probabilities, are also possible but not regarded here (see, e.g., Ben-Bassat 1978: 171 and DeGroot 1970: 431).

additionally gathered signal(s), while increased investigation costs more likely over-compensate these possible benefits (Mookerjee/Dos Santos 1993: 127).

2.3 Sequencing problem

Due to the assumption of a given sequence of criteria, the focus was on the optimal stopping of investigations, i.e. the minimization of expected total costs for a given sequence. However, next to solving the often nontrivial task of defining an adequate set of investigation criteria, the sequence of possible criteria matters to a great extent because it generally may change the realized posteriors and, therefore, expected total costs.¹⁵ This is due to the following facts:

The informational decision for a criterion at any stage cannot be seen isolated from previous informational decisions as well as consequential effects if the investigation is continued for at least one of the possible signal outcomes. Standard Bayesian updating (see (1)) reveals that for any given priors, different likelihoods lead to different posteriors, and for any given likelihoods, different priors will result in different posteriors. Thus, the (net) expected information value of any criterion is affected by its own position within the sequence as well as by the sequence of previous and subsequent criteria, which can be seen directly from (10) and (11). The example in Appendix B addresses the dependence of a criterion's (net) information value on the sequence of subsequent criteria, if the investigation is continued for at least one of the possible signal outcomes.

In this respect, Mussi (2002: 103) describes the investigation of a criterion as having the side-effect of changing the expected information value of other criteria. For example, the significance of certain reactions to a specific medical test depends on previously conducted tests and observed reactions; the investigation of initiation of resale price maintenance by manufacturers or retailers may provide different information about the anti-competitiveness if the market share of a specific practice has already been investigated compared to the case it has not. Furthermore, different criteria – when compared – can become advantageous for different priors (Hirshleifer/Riley 1992: 191) and, in general, the expected (net) information value of two or more criteria is not the sum of the single expected (net) information values (Kalagnanam/Henrion 1990: 272). Shortly stated, general principles concerning the estimation of information values are widely missing (Ferschl 1982: 71).

For these reasons, on the one hand, different sequences may not only result in different optimal stopping decisions but also in different realized expected total costs for a given class of cases; the minimization of total expected costs thus also requires the finding of the optimal sequence of criteria. On the other hand, finding the optimal sequence for one class of cases does not generally allow to draw conclusions regarding another one's optimal sequence.

¹⁵ Only if all possible criteria are stochastically independent and, in addition, the investigation will be continued until the set of possible criteria is exhausted for any combination of signal outcomes, the sequence of criteria does not matter to the realized probabilities of the states of the world. In the case of stochastically independent criteria, the joint likelihoods are simply the product of the single corresponding likelihoods (DeGroot 1962: 408).

The determination of the optimal sequence of a class of cases can generally be done by the examination of the entire decision tree, which encompasses every possible sequence of criteria, by backwards induction.¹⁶ Optimal rules then minimize the expected total costs at every point of the investigation process (Lewis/Sheehan 1990: 371; Vos 2007: 612). Unfortunately, the large number of possible sequences raises concerns about their calculation.

By splitting the investigation process into stages (due to the assumption of collecting only one signal at a time), we simultaneously defined each stage – except for the prior investigation stage – as consisting of at least two investigation points (depending on the realized outcomes at previous stages). An investigation path through the stages of a decision tree is therefore defined by at most one investigation point at every stage. For simplicity, we assume that every criterion can be investigated only once (cannot be repeated) on every such investigation path. Then, the following distinction has to be made (Kerber et al. 2008: 11-12):

1. *Stage-consistent sequences* mean that the same criterion is investigated at every investigation point of an investigation stage. The number of possible sequences is then given by the factorial of the possible number of criteria, i.e. $k!$; in the cases of $k = 3$, $k = 4$ and $k = 5$, e.g., the number of possible sequences are given by 6, 24 and 120, respectively.
2. *Flexible sequences* mean that a given criterion can be investigated at different stages for different investigation paths. The discrimination power of a criterion thus differs according to previously investigated criteria and observed signal outcomes (stochastically dependent likelihoods); for example, the investigation of the adoption-rate of resale price maintenance may provide different information for cases with low manufacturer market concentration compared to those with high manufacturer market concentration. Generally, the number of possible sequences is then given by the factorial of the number of possible criteria, with a power for every number that doubles for the decreasing order of them (beginning with one for the highest number), i.e. $k^{2^{k-k}} \cdot (k-1)^{2^{k-(k-1)}} \cdot (k-2)^{2^{k-(k-2)}} \cdot (k-3)^{2^{k-(k-3)}} \dots (k-(k-2))^{2^{k-(k-(k-2))}} \cdot (k-(k-1))^{2^{k-(k-(k-1))}}$; in the cases of $k = 3$, $k = 4$ and $k = 5$, e.g., the number of possible sequences are given by 12, 576 and 1,658,880, respectively.

The distinction makes clear that the number of possible sequences, i.e. the entire decision tree, is incredibly large (Hadley 1967: 153; Kalagnanam/Henrion 1990: 272) and – especially in the case of a flexible sequence – intractable without any computational aid. Unfortunately, no general solution exists with which the optimal sequence can be determined in advance by only regarding the criteria's likelihoods (Bradt/Karlin 1956: 390-391; Hadley 1967: 153; Heckerman et al. 1992: 97; Mussi 2002: 104). However, certain heuristics do exist, which try to approximate the optimal sequence in particular decision problems. While producing sub-optimal solutions in general, the heuristics' advantage of reducing a decision tree's size (i.e. the number of decision nodes) may offset this disadvantage (Gorry 1967: 39-40).

¹⁶ See DeGroot (1962: 410), Gorry (1967: 37-39), Gorry/Barnett (1968: 496-497), Heckerman et al. (1993: 292), Laux (1993: 118) or Mussi (2002: 103)

The sequencing problem in Bayesian sequential procedures can also be “solved” by determining a distribution function, which chooses over the next selected criteria (Mag 1977: 225). Such a random selection of sequences is done, for example, in psychological and educational questions with the aim of determining students’ level of knowledge (Lewis/Sheehan 1990, 1992; Vos 1999, 2007). Two distinctions are made: the number of tested items (in the above sense: criteria) may be fixed in advance (fixed-length testing), or the number of tested items out of a given set may vary (variable-length testing). In the first case, only two decisions are possible: passing or failing. The decision minimizes total expected losses and is revealed by (1) the number of correct answers if the tests do not differ in difficulty or (2) must be evaluated for every possible branch if the tests may differ in discrimination power, i.e. difficulty (Lewis/Sheehan 1990: 371-373). In the second case, a third decision alternative must be examined, i.e. the decision to continue testing. The design of this sequential procedure allows for the random choice of items in advance. Optimal decision rules are then derived by backward induction (Lewis/Sheehan 1990: 374-375; Vos 1999: 272-273; Vos 2007: 612). Another possibility is to choose the next tested item after each answer: either randomly (Lewis/Sheehan 1992: 68) or based on the amount of provided information at that point of investigation (Lewis/Sheehan 1990: 367-368; the authors briefly survey different approaches). The latter can be done with respect to any information measure (see Ben-Bassat 1978) and would be equivalent to the myopic approach of Gorry/Barnett (1968) if quantitative losses next to probabilities are included (see section 3.1). Although, it may be a good advice in some areas of application, it cannot be classified as a direct solution to the sequencing problem in general.

Therefore, we will focus on the examination of a number of possible heuristics in the following three sections, which try to approximate the optimal solutions.

3. Heuristic solutions

3.1 Myopic approach

Heuristic solutions try to approximate the resulting expected total costs of the optimal sequence and thus alleviate the problem of intractability of computing all possible sequences. As Mussi (2002: 104) mentions, many applications use the ‘myopic approach’ of Gorry/Barnett (1968).

This approach does not analyze a decision tree in its full dimension, but restricts the depth of the analysis by only considering sequences of criteria of limited length (Gorry/Barnett 1968: 498). The restriction may consider, for example, only the next one or two stages by asking if there exist single criteria or combinations of criteria for which the ‘myopic information value’ is higher than the raised investigation costs. Thereby, all further (possible) investigation stages are cut off and expected costs are calculated and compared under the assumption of terminating the investigation after the considered length of sequence. If more than one sequence reduces expected costs, the one with the highest reduction is chosen and the ‘myopic

question' is asked again at the next stage (with a reduced criteria set left because already investigated criteria cannot be chosen again); if not, the investigation is terminated.¹⁷ The myopic approach's repetition of the 'myopic question' is seen as being inconsistent with its own assumptions because the decisionmaker chooses the criteria under the assumption of immediate termination after observing the signal outcomes (Heckerman et al. 1993: 294).

When considering only one further criterion at any stage, the formal definition of the 'myopic information value' (\tilde{IV}) of criterion i at stage j is given by the reduction in expected losses, which results from the investigation of this one additional criterion, i.e.

$$(12) \quad \begin{aligned} \tilde{IV}_i(h_j) = & \min\{EL(h_j, d^+), EL(h_j, d^-)\} \\ & - \left[q_i(h_j) \min\{EL(h_{j+1}^+, d^+), EL(h_{j+1}^+, d^-)\} \right. \\ & \left. + (1 - q_i(h_j)) \min\{EL(h_{j+1}^-, d^+), EL(h_{j+1}^-, d^-)\} \right], \end{aligned}$$

with EL as defined in (3). Thus, the following maximization problem must be solved:

$$(13) \quad \max_{i \in K_j} \{ \tilde{IV}_i(h_j) - c_i \},$$

with K_j as the set of criteria not yet investigated at stage j .

While the original model requires the consideration of all costs at further investigation stages when deciding on whether to study the next criterion, the myopic approach requires the decisionmaker only to decide whether the solution to (13) is still positive. If not, no more criteria should be studied (Gorry/Barnett 1968: 501), which assumes the decision in favor of termination if the solution to (13) is zero. This also refers to a cost-effective investigation procedure because both the reduction in expected losses as well as the raised investigation costs is considered; in general, costly and well discriminating criteria may be less cost-effective compared to cheap and less discriminating criteria (Mookerjee/Dos Santos 1993: 116). Therefore, the myopic choice of a specific criterion may also regard the reduction of expected losses per unit of investigation costs.¹⁸

The myopic heuristic induces larger total expected costs as compared to the optimal sequence of criteria. On the one hand, this approach may induce a suboptimal sequence even for stochastically independent criteria (Appendix C exemplifies this shortcoming). On the other hand, stopping decisions may be handed down too early even when the sequence of criteria is correct because further cost reductions resulting from improvements at later investigation stages are neglected (Kerber et al. 2008: 12-13). Both reasons of sub-optimality of the myopic approach are aggravated if the criteria are assumed to be not stochastically independent.

¹⁷ See, for example, Heckerman et al. (1992: 97), Kalaganam/Henrion (1990: 272), Mussi (2002: 104) and Tseng/Gmytrasiewicz (1999:2-4).

¹⁸ This is the way Mookerjee/Dos Santos (1993), Mookerjee/Mannino (1997) and Mannino/Mookerjee (1999) describe an inductive algorithm, which seeks to maximize the value of an expert system (defined by solving the problem of optimally choosing the next information input under the given trade-offs between better information and acquiring costs), i.e. maximizing the difference of the expected decision value and costs of collecting information inputs. Gorry/Barnett (1968: 501) explicitly allow for such a method used to develop a decision tree "without altering the basic method of the analysis".

Furthermore, investigation processes cannot be terminated too late by the myopic approach: should the complete study of possible sequences indicate that termination is the best alternative, then there obviously is no more criterion worthwhile to be investigated at the next stage; thus, the myopic criterion will also recommend termination.

Generally, by restricting the analysis to a sequence of limited length, the decisionmaker may overlook an effective sequence characterized by at least one more criterion (Gorry/Barnett 1968: 498). Sometimes the expected information value of a sequence of criteria is larger than the sum of their investigation costs, while this does not hold for single (respectively: a shorter sequence of) criteria (Mussi 2002: 103-104). Despite the obvious shortcomings, as Gorry/Barnett (1968: 505) argue, the myopic approach may be seen as a good heuristic in practice – which also refers to the possible offsetting advantage of a reduced decision tree's size (Gorry 1967: 39-40).¹⁹

3.2 'Bracketing condition'

The myopic approach raises the question if the myopic information value must be calculated at each stage for every remaining criterion. When regarding a sequence of only one criterion to be considered next, the determination of a superior criterion i , which dominates other criteria i' no matter which loss function and prior probabilities are presumed, would reduce the set of criteria coming into question.²⁰ For any given decision situation, such a superior criterion is characterized by lower total expected losses (i.e. a higher information value) compared to other possible criteria (Bradt/Karlin 1956: 391; Hirshleifer/Riley 1979: 1397; Ponssard 1975: 452). This condition is fulfilled if the posteriors of the superior criterion *bracket* those of an inferior; therefore the decisionmaker has a smaller posterior risk of making the wrong decision, no matter which outcome is observed (Hirshleifer/Riley 1979: 1397; Marschak/Miyasawa 1968: 144 et seqq.; Ponssard 1975: 452).²¹ Formally, we say that a criterion i 'brackets' another criterion $i' \neq i$ for a history h_j of collected signals and their outcomes if

$$(14) \quad \hat{p}(h_j \cap (i, +1)) > \hat{p}(h_j \cap (i', +1)) > \hat{p}(h_j \cap (i', -1)) > \hat{p}(h_j \cap (i, -1))$$

is satisfied, where $h_j \cap (i, m)$ denotes the extension of history h_j by the collection of signal i and its outcome m and $h_j \cap (i', m)$ is defined accordingly for signal i' . Without loss of generality it is presumed that $m = +1$ ($m = -1$) is more conclusive as to the occurrence of

¹⁹ Gorry/Barnett (1968) concretely consider the diagnosis of congestive heart disease. Their decision-theoretic approach can nevertheless be generalized to allow the conclusion also for other sequential decision problems.

²⁰ Such a determination is more difficult, or impossible in general, for a sequence of more than one criterion because the expected net information value of two or more criteria is not the sum of the single expected net information values (Kalagnanam/Henrion 1990: 272).

²¹ Davis (1994) evaluates different legal procedures (criteria) in resolving legal disputes. The legal procedures reveal the value of a continuous variable and can be ranked according to the fulfilment of a mean-preserving spread of the corresponding probabilities of guilt. This also describes better discrimination and concludes smaller decision error costs.

$w = B$ ($w = H$); otherwise the inequality chain will reverse. Since the condition only applies to the information (and not *net* information) value (Marschak/Miyasawa 1968: 148), its application is limited to cases of equally costly criteria and cases where investigation costs of an inferior criterion are at least as high as those of a superior criterion. The proof of lower total expected losses is given in Appendix D.

The ‘bracketing condition’ is revealed at the same time also by the direct characteristics of the criteria, which are given by the likelihoods: when those of an inferior criterion are obtained from those of the superior criterion by a stochastic transformation, the former is less informative for any priors and loss functions, and thus always reveals higher expected losses.²² Thus, the superior criterion is characterized by a higher discriminatory power of the possible signal outcomes, which means that the signal outcomes can more probably be assigned to the true state of the world. In the case of $m = +1$ ($m = -1$) being more likely for $w = B$ ($w = H$), we get $q_i^B > q_i^H$ and $q_i^H < q_i^B$ (Appendix D also provides the proof of lower total expected losses in this case).

The bracketing condition is satisfied for pairs of criteria, leading to a ranking of the fulfilling criteria. This ranking holds for any priors and loss functions due to the equivalent expressions either in terms of the resulting posteriors or in terms of the underlying likelihoods (Ponssard 1975: 449-452). If, on the other hand, the bracketing condition is not fulfilled for some pair of criteria, which thereof is more informative depends on the specifics of the decision situation (Hirshleifer/Riley 1979: 1397). Thus, the bracketing condition is a strongly sufficient condition for more informativeness (Hirshleifer/Riley 1992: 198).²³ Difficulties may arise in the case of stochastic dependency among the outcomes of the criteria out of K . In that case, to name one criterion superior to another requires that the bracketing condition is fulfilled for any history of signals and their outcomes. Nevertheless, even in case of stochastic dependency the bracketing condition can reduce the number of criteria to be considered at each specific stage. To summarize, the fulfillment of the bracketing condition simplifies the myopic approach’s application.

Given that the myopic approach, in general, does not lead to the optimal solution, it may be asked whether this can still be accomplished under (special) conditions. At first glance, the bracketing condition might be a candidate for the general ex ante determination of optimal investigation solutions. Unfortunately, the bracketing condition cannot be applied to yield a general optimal sequence. Although it is optimal in a one-step decision problem to always choose the superior criterion, the analogy for multi-stage decision problems is not true in general. Even for a set of stochastically independent and equally costly criteria, which are completely ordered due to the pair-wise fulfillment of the bracketing condition, it is not generally optimal to always investigate the criteria in a decreasing order, i.e. beginning with

²² See, e.g., Athey/Levin (2001: 2-3), Bielinska-Kwapisz (2003: 21 et seq.), Blackwell (1951, 1953), Blackwell/Girshick (1954: 324-336), Crémer (1982: 439 et seq.), Hirshleifer/Riley (1992: 187-193), Marschak (1964: 53), Marschak/Miyasawa (1968: 144 et seq.) or Ponssard (1975: 449-452).

²³ DeGroot (1962) and Ponssard (1975) regard the optimal design of sequential experiments (in the above framework: investigation of criteria). In the case of independent experiments and the existence of a superior one, the optimal investigation rule – the one that minimizes expected total loss – requires the replication of the superior experiment (DeGroot 1962: 412; Ponssard 1975: 452).

the most superior criterion (the one that brackets all other criteria) and ending with the most inferior criterion (the one that is bracketed by all other criteria). Stated differently, sometimes expected total costs are reduced by more at stage $j+1$ if an inferior probability distribution – left by a worse criterion at stage j – is discriminated better than if a superior probability distribution – left by a better criterion at stage j – is discriminated worse. Intuitively, it sometimes pays off to exercise strong criteria on remaining uncertain subclasses at later stages of the investigation instead of applying the same criteria earlier without the additional benefit of shortening some paths of the whole investigation process. Therefore, the bracketing condition can only be applied to the myopic approach if one criterion is considered.

3.3 More heuristics

Heckerman et al. (1993) consider so-called non-myopic approximations for the special case of dichotomy among states of the world, decision possibilities and signal outcomes.²⁴ Each signal is conditionally independent of the observation of other signals' outcomes. In the case of independent investigation costs of all possible criteria, the approximation works as follows: given that no single criterion is worthwhile investigating at any stage according to the myopic approach, the remaining criteria should be arranged in descending order of their net information value; at next, all subsequences starting with the most (net) informative criterion are regarded and the smallest subsequence with a positive net information value is chosen (Heckerman et al. 1993: 298). However, the myopic approach of Gorry and Barnett explicitly allows for the consideration of sequences of criteria (Gorry/Barnett 1968: 498) and does not automatically mean to consider the net information value of only a single criterion at any stage. Therefore, their approach might be classified as more general than the one of Heckerman et al. (1993).

Another interesting special analysis of Heckerman et al. (1993) regards dependencies among the investigation costs of some criteria. The examination of a sequence of criteria (of a defined length) must consider all possible combinations of the outcomes. Thus, there are combinations leading to the optimality of one decision possibility and combinations leading to the optimality of the opposite decision possibility. Total expected costs of both decisions must then include the expectations over all corresponding combinations. Since criteria can only be performed as a set, all net information value analysis must be applied analogously.

Gorry (1967) mentions a few more heuristics, which are more or less applicable to the above framework. In general, he distinguishes between so-called breadth-limiting and depth-limiting heuristics (ibid.: 43): the first limit the number of analyzed criteria at each stage, and the second limit the number of stages looked ahead. The myopic approach of section 3.1 must, therefore, be classified as a depth-limiting heuristic (Gorry/Barnett 1968: 498).

²⁴ Heckerman et al. (1993: 293) use the notions of hypotheses, action (or alternatives), and instantiations of pieces of evidence (or tests).

Applying a breadth-limiting heuristic, the decisionmaker may restrict attention to some fixed number of criteria at each stage (Gorry 1967: 43-44). The main problem is the specification of a selection rule because it always requires the examination of information values, which include effects on subsequent stages. The heuristic's purpose (of limiting the considered breadth) is thus self-destroyed. Another breadth-limiting heuristic, therefore, ignores effects on information values and the information embodied in prior probabilities (ibid.: 44-45). At each stage the number of possible criteria is limited and chosen according to the increased order of their investigation costs. When regarding more than two possible states of the world, the selection may comprise only criteria with a (stronger) relationship to the most probable states; which, on the other hand, does not necessarily mean to exclude any criterion (ibid.: 45-46). However, such selection rule – based purely on current probability distributions – somehow risks driving the subsequent investigation process in a predetermined direction.

Depth-limiting heuristics can be either characterized by fixed depth to be considered at each stage or by variable depth; the latter try to take into account some kind of 'information potential' of different probabilities at each stage (ibid.: 48). In fact, this is the crucial problem of finding an optimal solution to sequential decision problems and thus supports the sub-optimality of heuristics. As mentioned in the previous section, depth-limiting heuristics suffer from neglecting further cost reductions at later stages and thus indicate termination at some points, at which the optimal solution requires further investigation (ibid.: 49-50).

These additional heuristics, however, have to be questioned in regard to their actual purpose of reducing the computational effort. In result, the myopic solution (with short sequences of criteria to be regarded) seems to be the most promising approach; especially in conjunction with the simplifying bracketing condition. However, the above-listed heuristics become more important for special situations. For example, knowledge about dependencies among the investigation costs of some criteria would make it necessary to somehow account for the second heuristic of Heckerman et al. (1993); application in combination with another heuristic may create a new approach. Furthermore, knowledge about incompatibilities among some investigation criteria functions like a selection rule for breadth-limiting heuristics.

Despite the non-existence of an optimal approach as well as a best heuristic, predictions about the recommendation of one or another heuristic may be possible in the case of incomplete knowledge about certain characteristics of the investigation criteria. For example, choosing among the criteria according to the increased order of their investigation costs may be preferable or the only option if too little is known about the discrimination characteristics of the available criteria. On the other hand, a lack of knowledge concerning the associated investigation costs may necessitate assuming equal costs, which lets the myopic approach look reasonable. However, it is difficult to be more precise concerning a general preference among the heuristics; one cannot help but examine the available heuristics in concrete decision situations with specific constellations and knowledge about the investigation criteria's characteristics.

4. Discussion and conclusions

A lot of different areas in ordinary and professional life can be characterized as statistical decision problems, in which particular cases are assessed corresponding to their true and ex ante unknown state of the world. Classifying different decision situations subject to recurring or corresponding characteristics provides decisionmakers with information about respective probabilities of the possible states of the world. The subsequent assessment follows a sequential investigation of specific features (so-called investigation criteria), which leads to a refinement of initially presumed probability distributions, according to the observation of respective realizations that are linked to the possible states of the world. This procedure sorts particular cases into different and ever smaller subclasses.

For example, reacting to certain medical tests in specified ways (or showing specific symptoms), leads to a refinement of the probabilities of specific diseases (Gorry 1967: 23; Warner/Veasey 1992: 43 et seqq.). Answering to certain examination questions in a correct way increases the probability of being examined as a high-knowledge student; vice versa for incorrect answers (Vos 1999: 273; 2007: 617). The assessment of market indicators (e.g. market concentration, buyer power or existing R&D agreements) refines the probabilities of pro- and anticompetitive welfare effects of specific business behaviors and thus the potential of exerting market power.

The comparison of obtained information and raised costs may indicate an optimal termination prior to the investigation of every possible criterion (or test) out of the relevant set. This optimal stopping problem is one of the crucial characteristics of sequential investigations.²⁵ Another one is a generally unsolved problem: namely, the optimal sequencing of investigation criteria. As mentioned above, different sequences affect the information gained by investigating the criteria and thus result in different expected total costs (i.e. the sum of expected losses through wrong decisions and the costs of investigating the criteria).

Despite its obvious theoretical significance, one may question how substantial the sequencing problem is in specific decision situations. This cannot be generalized, but certainly depends on the distribution of losses and investigation costs of the different criteria. Furthermore, the relevance of the problem increases with the depth of realized investigation paths. On the other hand, the prior probability distribution does not seem to be a promising indicator in this regard; depending on the set of available criteria, relatively unambiguous classes often require numerous criteria to be investigated in specific paths, while it suffices to roughly subdivide relatively 'uncertain' classes. However, the sequencing problem becomes especially important if different sequences lead to (substantially) different investigation structures, i.e. different stopping points within the analogously same investigation paths.

Even though the optimal solution is derived by backward induction, this remains tractable only for decision problems characterized by a manageable number of criteria. Unfortunately, no general solution exists with which the optimal sequence can be determined in advance (i.e. by purely considering the criteria's characteristics). Therefore, this paper analyzes existing

²⁵ Of course, diagnostic decision problems first require the definition of a set of relevant criteria.

heuristics trying to circumvent the sequencing problem by approximating the optimal solution and its expected total costs.

The approximation by means of myopic approaches probably proves to be the superior method in practice, at least for the majority of decision situations requiring the subclassification of behaviors or actions. The main advantage lies in fast sorting of cases into clearly differentiated subclasses. This is secured by investigating the most powerful criteria first and this is what the solution of many practical decision problems often requires. However, preference for one (heuristic) solution over another crucially depends on the specific characteristics of concrete decision situations. Incomplete knowledge about some characteristics or constellations might also require adjustments towards one or another heuristic.

Unfortunately, the ubiquity of statistical decision problems in lots of different areas lacks the non-existence of a general solution to the sequencing problem. Despite their usefulness, the available working approximations cannot hide this fact. One possible approach for future research may be the analysis of specific decision situations narrowed by (simplifying) assumptions about certain characteristics and constellations. Present research results cannot exclude the derivation of optimal solutions for subsets of sequential decision problems either.

Appendix A

This appendix proves that the reversion of Bayesian updating only holds for the general discrete case if the number of possible states of the world at least equals the number of possible signal outcomes. The possible states of the world $w=1, \dots, n$ (basic model case: $w \in \{B, H\}$) are characterized by prior probabilities $p(w)$ ('priors'). Every criterion has the same number of possible signal outcomes $m=1, \dots, s$ (basic model case: $m \in \{-1, +1\}$), which are characterized by the likelihoods $q(m|w)$. Then, if any one of the outcomes m is observed, standard Bayesian updating yields:

$$\hat{p}(w|m) = \frac{q(m|w)p(w)}{\sum_{w=1}^n q(m|w)p(w)},$$

with $\sum_{w=1}^n q(m|w)p(w) = q(m)$ as the unconditional probability of observing m and $\hat{p}(w|m)$ as the posterior probability of the true state of the world w .

Now, with s possible signal outcomes we get s posterior probabilities for each state of the world, i.e. we get s independent equations for each state of the world (illustrated by the columns in the figure below). Given any outcome m , the posteriors for the true state of the world w add to one, i.e. we get $n-1$ independent equations for each signal outcome

(illustrated by the rows in the figure below). Thus, in total we have $s \times (n-1)$ independent equations.

	$w=1$	\dots	$w=n$	
$m=1$	$\hat{p}(w=1 m=1) = \frac{q(m=1 w=1)p(w=1)}{q(m=1)}$	\dots	$\hat{p}(w=n m=1) = \frac{q(m=1 w=n)p(w=n)}{q(m=1)}$	$\sum_w = 1$
\vdots	\vdots	\vdots	\vdots	\vdots
$m=s$	$\hat{p}(w=1 m=s) = \frac{q(m=s w=1)p(w=1)}{q(m=s)}$	\dots	$\hat{p}(w=n m=s) = \frac{q(m=s w=n)p(w=n)}{q(m=s)}$	$\sum_w = 1$

To calculate consistent conditional probabilities with only the knowledge of the priors and posteriors, we need $s-1$ equations for each of the n states of the world (because the likelihoods add to one for every state of the world w), i.e. we need $(s-1) \times n$ equations.

	$w=1$	\dots	$w=n$
$m=1$	$q(m=1 w=1)$	\dots	$q(m=1 w=n)$
\vdots	\vdots	\vdots	\vdots
$m=s$	$q(m=s w=1)$	\dots	$\hat{p}(m=s w=n)$
	$\sum_m = 1$	\dots	$\sum_m = 1$

It is only possible to calculate consistent likelihoods if the number of independent equations at least equals the number of needed equations, which holds if:

$$s \times (n-1) \geq (s-1) \times n \Leftrightarrow n \geq s.$$

The same result can be derived directly from the unconditional probabilities via the consistency conditions for prior and posterior probabilities: $p(w) = \sum_{m=1}^s q(m) \hat{p}(w|m)$. For n possible states of the world we have $n-1$ independent consistency conditions. Because the unconditional probabilities add to one ($\sum_{m=1}^s q(m) = 1$), we need $s-1$ equations.

$$\begin{array}{ccccccc}
& & m=1 & & \dots & & m=s \\
\hline
p(w=1) & = & q(m=1)\hat{p}(w=1|m=1) & + & \dots & + & q(m=s)\hat{p}(w=1|m=s) \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot & & \cdot \\
\hline
p(w=n) & = & q(m=1)\hat{p}(w=n|m=1) & + & \dots & + & q(m=s)\hat{p}(w=n|m=s) \\
\hline
\sum_{w=1}
\end{array}$$

To summarize, if the number of possible states of the world is less than the number of possible signal outcomes, we are unable to express the likelihoods and unconditional probabilities only in terms of priors and posteriors of the states of the world; intuitively, more than one signal outcome has to be assigned to at least one of the possible states of the world, which produces inconclusive information.²⁶

Appendix B

Consider the following simple example with three stochastically independent criteria ($K = 3$). The example addresses the dependence of a criterion's (net) information value on the sequence of subsequent criteria if the investigation is continued for at least one of the possible signal outcomes. Losses are assumed to be symmetric and normalized to one. Investigation costs are the same for all criteria and amount to 0.04 times the losses ($c = 0.04$).

	q_i^B	q_i^H
$i = 1$	0.7	0.28
$i = 2$	0.74	0.3
$i = 3$	0.75	0.35

Next, consider the class of cases characterized by $p = 0.35$ and the two possible sequences starting with the second criterion $i = 2$, i.e. continuing with $i = 1$ or $i = 3$ first: both sequences are terminated in the cases of observing $m = -1$ at the first stage as well as observing $m = +1$ at the first and second stage, and completely investigated when observing $m = +1$ at the first and $m = -1$ at the second stage. The net information value of criterion $i = 2$ is defined by equation (11):

²⁶ Hirshleifer/Riley (1992: 196-197) provide the same claim, however, without a proof.

$$NIV_2(h_0) = EC^*(h_0) - \left[q_2(h_0) EC^*(h_1^{+1}) + (1 - q_2(h_0)) EC^*(h_1^{-1}) \right].$$

First, expected total costs of immediate termination, $EC^*(h_0)$, are given by the lower prior probability $p = 0.35$ and thus independent of the subsequent sequence.

Second, expected total costs are the same for both sequences when observing $m = -1$ at the first stage, because both start with the same criterion: $EC^*(h_1^{-1}) = c + \hat{p}(h_1^{-1}) = 0.20\bar{6}$. From equations (5) we know that the unconditional probabilities of the first signals can be expressed in terms of the priors and first stage posterior probabilities. Consequently, $q_2(h_0)$

is also the same for both sequences: $q_2(h_0) \stackrel{(2)}{=} 0.35 \cdot 0.74 + 0.65 \cdot 0.3 = 0.454$.

Third, when observing $m = +1$ at the first stage, expected total costs, $EC^*(h_1^{+1})$, are reduced by a larger amount if continued with $i = 3$ first, instead of $i = 1$. The increased discrimination power of the former sequence is given by

$$EC^*(h_1^{+1}) = q_3(h_1) EC(h_2, d^+) + (1 - q_3(h_1)) EC(h_2, d^0) = 0.3682$$

compared to

$$EC^*(h_1^{+1}) = q_1(h_1) EC(h_2, d^+) + (1 - q_1(h_1)) EC(h_2, d^0) = 0.3705.$$

Figures 1 and 2 summarize the second-stage calculations for both sequences by means of a simple decision tree illustration, which is based on Kerber et al. (2008: 6). To summarize, the net information value of $i = 2$ is given by $NIV_2(h_0) = 0.35 - [0.454 \cdot 0.36819 + 0.546 \cdot 0.20\bar{6}] = 0.07$ if continued with $i = 3$ first and by $NIV_2(h_0) = 0.35 - [0.454 \cdot 0.3705 + 0.546 \cdot 0.20\bar{6}] = 0.068953$ if continued with $i = 1$ first.

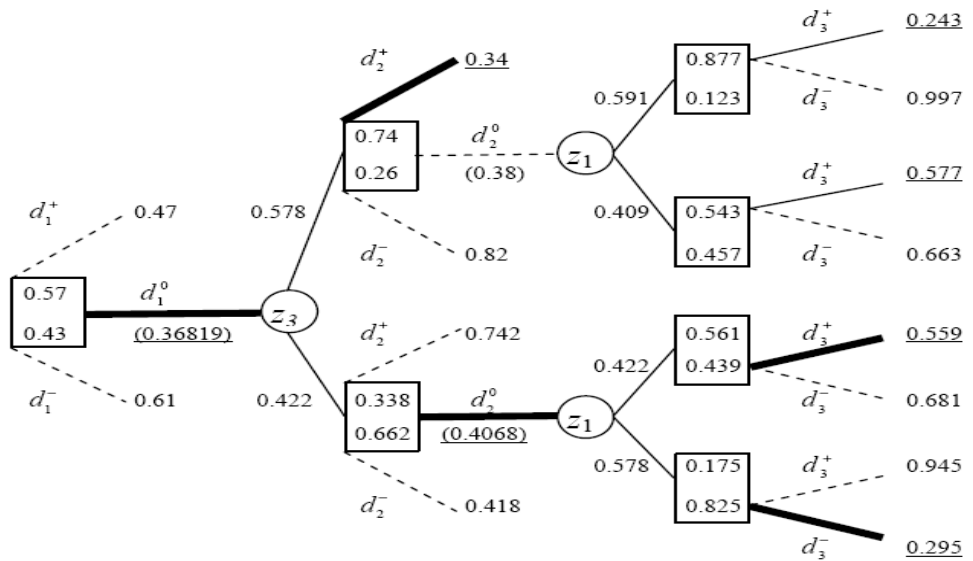


Figure 1: Second-stage investigation of $i = 3$.

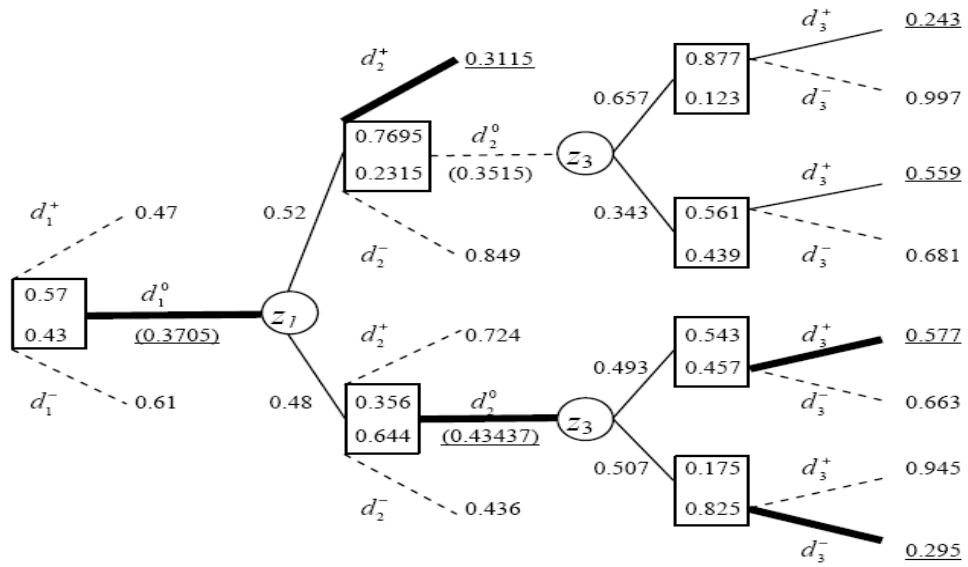


Figure 2: Second-stage investigation of $i = 1$.

Appendix C

The following example addresses the myopic approach's failing in finding the optimal sequence of criteria. Reducing the calculation to three possible criteria ($K = 3$) and a stage-consistent sequence, the examination focuses on the second stage.

	q_i^B	q_i^H
$i = 1$	0.82	0.18
$i = 2$	0.72	0.41
$i = 3$	0.68	0.34

The optimal decision tree for $p = 0.24$ is given in figure 3. When we rely on the myopic approach (regarding a sequence of only one criterion) to determine the sequence of criteria, the information value of criterion $i = 2$ is underestimated (see figures 5 and 6 for the second-stage calculations of expected costs according to the myopic approach), since for one outcome the investigation does not terminate but calls for criterion $i = 3$. As may be seen from the myopic decision tree in figure 4, the same is true for criterion $i = 3$ if investigated at the second stage; but here the effect is weaker (see figures 5 and 6). As a consequence, the sequence of these two criteria is reversed.

Assuming that losses are symmetric and the investigation costs are the same for all criteria ($c = 0.04$), the expected total costs of the optimal decision tree are given by 0.2031 and those of the myopic decision tree by 0.2038.

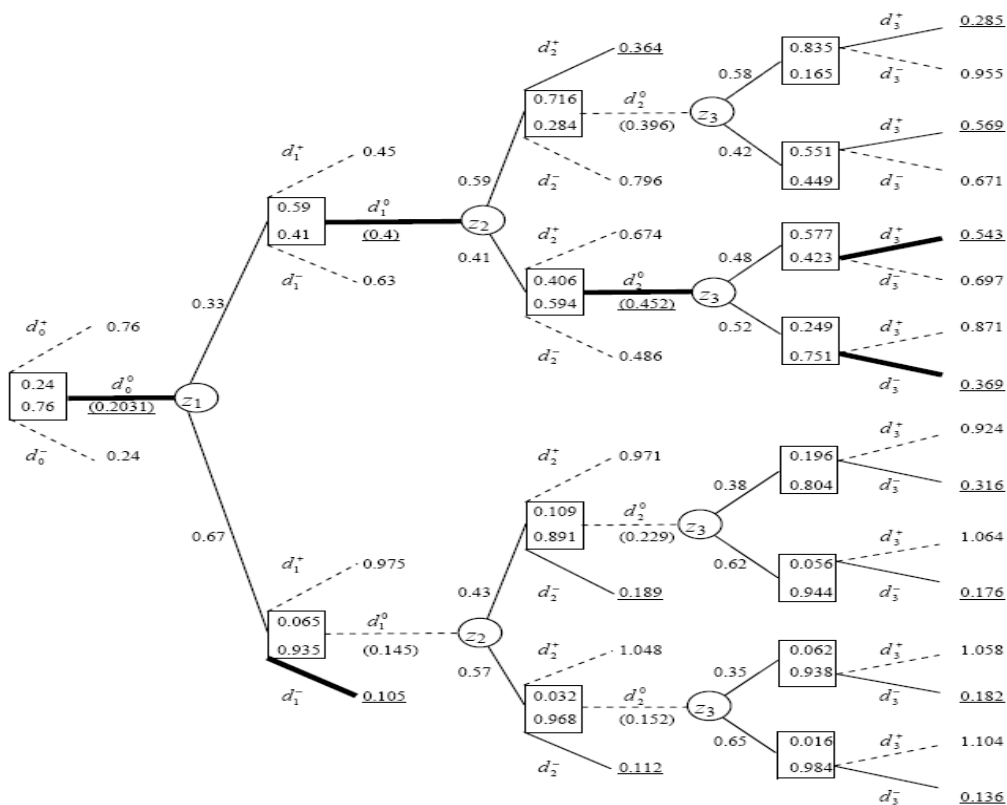


Figure 3: Optimal decision tree

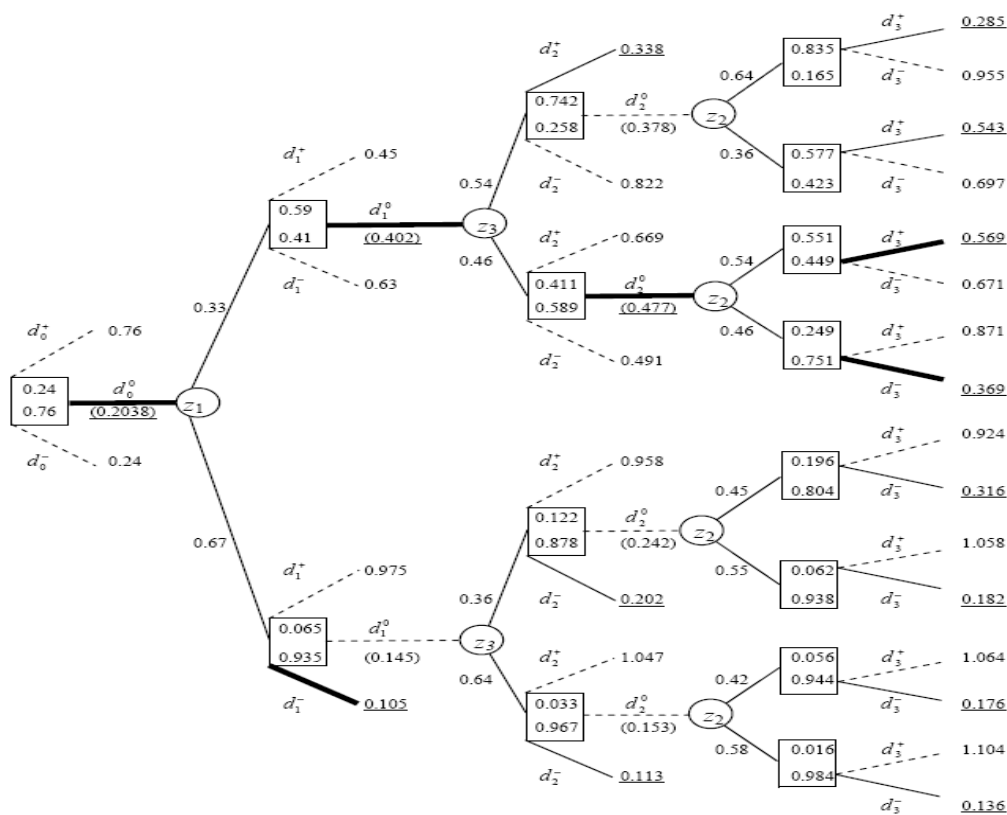


Figure 4: Myopic decision tree

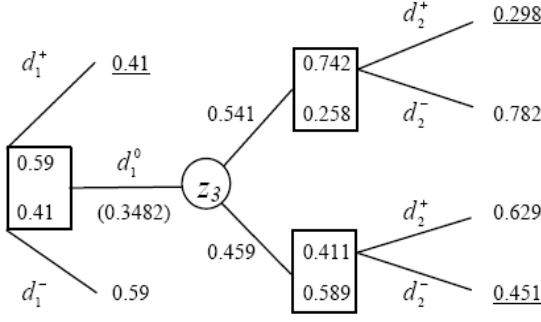


Figure 5: Second-stage criterion according to the myopic approach

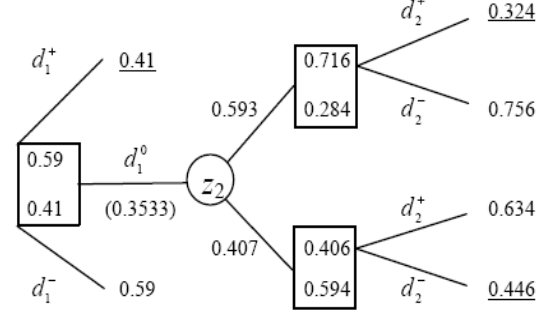


Figure 6: Optimal second-stage criterion

Appendix D

This appendix proves lower total expected losses for any given priors and loss function if a superior criterion i brackets another possible criterion i' out of K . The lower total expected losses are equivalent with a higher reduction of total expected losses, which indicates more informativeness. For this to hold, the posteriors of criterion i' are derived by a stochastic transformation of the posteriors of another criterion i , wherefore the posteriors of w of criterion i' can be expressed as a convex combination of the posteriors of w of criterion i , regarding both signals. This proof of more informativeness is known as *Blackwell's Theorem*.²⁷ The first part derives conditions, which hold for the unconditional signal-probabilities if the bracketing condition is fulfilled; the second part then proves the resulting lower total expected losses of criterion i .

Starting point of the analysis is any stage j in the investigation process with the given history of signals h_j . To simplify the expressions, we abbreviate the notation as follows: $\hat{p}(h_j \cap (i, m))$ is shortly written as $\hat{p}_i(m)$ and q_i abbreviates $q_i(h_j)$. Obviously, the expected loss may only be reduced by one-step investigation if the terminating decision depends on the signal outcome for at least one of the two criteria. Assuming without loss of generality that the outcome $m = +1$ ($m = -1$) is more conclusive as to the occurrence of $w = B$ ($w = H$), this implies the distinction of five different cases:

(i) the terminating decision depends on the signal's outcome for both criteria,

$$(A.1) \quad \hat{p}_i(+1) > \hat{p}_{i'}(+1) \geq \frac{L_H}{L_H + L_B} > \hat{p}_{i'}(-1) > \hat{p}_i(-1);$$

(ii) the terminating decision depends on the signal's outcome for one of the two criteria, which can by definition of the bracketing condition (14) only be criterion i ,

²⁷ See Blackwell (1953), Blackwell/Girshick (1966: 324 et seqq.) and Hirshleifer/Riley (1992: 188 et seqq.)

$$(A.2) \quad \hat{p}_i(+1) \geq \frac{L_H}{L_H + L_B} > \hat{p}_{i'}(+1) > \hat{p}_{i'}(-1) > \hat{p}_i(-1), \text{ or}$$

$$(A.3) \quad \hat{p}_i(+1) > \hat{p}_{i'}(+1) > \hat{p}_{i'}(-1) \geq \frac{L_H}{L_H + L_B} > \hat{p}_i(-1);$$

(iii) the terminating decision is equal for both signal outcomes,

$$(A.4) \quad \hat{p}_i(+1) > \hat{p}_{i'}(+1) > \hat{p}_{i'}(-1) > \hat{p}_i(-1) \geq \frac{L_H}{L_H + L_B}, \text{ or}$$

$$(A.5) \quad \frac{L_H}{L_H + L_B} \geq \hat{p}_i(+1) > \hat{p}_{i'}(+1) > \hat{p}_{i'}(-1) > \hat{p}_i(-1).$$

Both cases of (iii) do not fulfill the requirement that the terminating decision depends on the signal's outcome for at least one of the two criteria. Thus, both criteria are characterized by a zero information value in these cases. The analysis follows Hirshleifer/Riley (1992: 187-193).

First, the expression of the posteriors of criterion i' as a convex combination of the posteriors of criterion i ('bracketing condition'), is given by:

$$(A.6) \quad \hat{p}_i(+1)a_+ + \hat{p}_i(-1)(1-a_+) = \hat{p}_{i'}(+1) \quad \text{and} \quad \hat{p}_i(+1)a_- + \hat{p}_i(-1)(1-a_-) = \hat{p}_{i'}(-1).$$

The prior probabilities of the true state of the world at any stage j can always be expressed as the average of the corresponding posteriors of any criterion at the next stage, weighted by the unconditional probabilities of observing the two signal outcomes, i.e.

$$(A.7) \quad \hat{p}_i(+1)q_i + \hat{p}_i(-1)(1-q_i) = \hat{p}(h_j) = \hat{p}_{i'}(+1)q_{i'} + \hat{p}_{i'}(-1)(1-q_{i'}).$$

Inserting (A.6) yields

$$(A.8) \quad \begin{aligned} & \hat{p}_i(+1)q_i + \hat{p}_i(-1)(1-q_i) \\ &= (\hat{p}_i(+1)a_+ + \hat{p}_i(-1)(1-a_+))q_{i'} + (\hat{p}_i(+1)a_- + \hat{p}_i(-1)(1-a_-))(1-q_{i'}), \end{aligned}$$

which simplifies to:

$$(A.9) \quad q_i = a_+q_{i'} + a_-(1-q_{i'}).$$

Now, total expected loss of i' , $EL_{i'}^*(h_j, d^0) = q_{i'} \cdot EL_{i'}^*(+1) + (1-q_{i'}) \cdot EL_{i'}^*(-1)$, expands to:

$$(i) \quad EL_{i'}(h_j, d^0) = q_{i'} \cdot L_H (1 - \hat{p}_{i'}(+1)) + (1 - q_{i'}) \cdot L_B \hat{p}_{i'}(-1)$$

$$\begin{aligned} & \stackrel{(B.6)}{=} q_{i'} L_H ((1 - \hat{p}_i(+1))a_+ + (1 - \hat{p}_i(-1))(1 - a_+)) + (1 - q_{i'}) L_B (\hat{p}_i(+1)a_- + \hat{p}_i(-1)(1 - a_-)) \\ &= q_{i'} (a_+ (1 - \hat{p}_i(+1)) L_H + (1 - a_+) (1 - \hat{p}_i(-1)) L_H) + (1 - q_{i'}) (a_- \hat{p}_i(+1) L_B + (1 - a_-) \hat{p}_i(-1) L_B) \end{aligned}$$

$$\begin{aligned}
& \stackrel{(B.1)}{>} q_i \left(a_+ (1 - \hat{p}_i(+1)) L_H + (1 - a_+) \hat{p}_i(-1) L_B \right) + (1 - q_i) \left(a_- (1 - \hat{p}_i(+1)) L_H + (1 - a_-) \hat{p}_i(-1) L_B \right) \\
& = (q_i a_+ + (1 - q_i) a_-) (1 - \hat{p}_i(+1)) L_H + (q_i + 1 - q_i - q_i a_+ - (1 - q_i) a_-) \hat{p}_i(-1) L_B \\
& \stackrel{(B.9)}{=} q_i (1 - \hat{p}_i(+1)) L_H + (1 - q_i) \hat{p}_i(-1) L_B \\
& = EL_i(h_j, d^0),
\end{aligned}$$

$$(ii) \quad EL_{i'}(h_j, d^0) = q_{i'} \cdot L_B \hat{p}_{i'}(+1) + (1 - q_{i'}) \cdot L_B \hat{p}_{i'}(-1)$$

$$\begin{aligned}
& \stackrel{(B.6)}{=} q_{i'} L_B (\hat{p}_i(+1) a_+ + \hat{p}_i(-1) (1 - a_+)) + (1 - q_{i'}) L_B (\hat{p}_i(+1) a_- + \hat{p}_i(-1) (1 - a_-)) \\
& = q_{i'} (a_+ \hat{p}_i(+1) L_B + (1 - a_+) \hat{p}_i(-1) L_B) + (1 - q_{i'}) (a_- \hat{p}_i(+1) L_B + (1 - a_-) \hat{p}_i(-1) L_B) \\
& \stackrel{(B.2)}{>} q_{i'} \left(a_+ (1 - \hat{p}_i(+1)) L_H + (1 - a_+) \hat{p}_i(-1) L_B \right) + (1 - q_{i'}) \left(a_- (1 - \hat{p}_i(+1)) L_H + (1 - a_-) \hat{p}_i(-1) L_B \right) \\
& = (q_{i'} a_+ + (1 - q_{i'}) a_-) (1 - \hat{p}_i(+1)) L_H + (q_{i'} + 1 - q_{i'} - q_{i'} a_+ - (1 - q_{i'}) a_-) \hat{p}_i(-1) L_B \\
& \stackrel{(B.9)}{=} q_{i'} (1 - \hat{p}_i(+1)) L_H + (1 - q_{i'}) \hat{p}_i(-1) L_B \\
& = EL_i(h_j, d^0)
\end{aligned}$$

$$\text{or} \quad EL_{i'}(h_j, d^0) = q_{i'} \cdot L_H (1 - \hat{p}_{i'}(+1)) + (1 - q_{i'}) \cdot L_H (1 - \hat{p}_{i'}(-1))$$

$$\begin{aligned}
& \stackrel{(B.6)}{=} q_{i'} L_H \left((1 - \hat{p}_i(+1)) a_+ + (1 - \hat{p}_i(-1)) (1 - a_+) \right) + (1 - q_{i'}) L_H \left((1 - \hat{p}_i(+1)) a_- + (1 - \hat{p}_i(-1)) (1 - a_-) \right) \\
& = q_{i'} \left((1 - \hat{p}_i(+1)) a_+ L_H + (1 - \hat{p}_i(-1)) (1 - a_+) L_H \right) + (1 - q_{i'}) \left((1 - \hat{p}_i(+1)) a_- L_H + (1 - \hat{p}_i(-1)) (1 - a_-) L_H \right) \\
& \stackrel{(B.3)}{>} q_{i'} \left((1 - \hat{p}_i(+1)) a_+ L_H + \hat{p}_i(-1) (1 - a_+) L_B \right) + (1 - q_{i'}) \left((1 - \hat{p}_i(+1)) a_- L_H + \hat{p}_i(-1) (1 - a_-) L_B \right) \\
& = (a_+ q_{i'} + a_- (1 - q_{i'})) (1 - \hat{p}_i(+1)) L_H + (1 - a_+ q_{i'} - a_- (1 - q_{i'})) \hat{p}_i(-1) L_B \\
& \stackrel{(B.9)}{=} q_{i'} (1 - \hat{p}_i(+1)) L_H + (1 - q_{i'}) \hat{p}_i(-1) L_B \\
& = EL_i(h_j, d^0).
\end{aligned}$$

Therefore, if the posteriors of any two criteria i and i' fulfill the ‘bracketing condition’, the superior criterion i always results in lower total expected losses ($EL_i(h_j, d^0) < EL_{i'}(h_j, d^0)$) with respect to any loss function and prior probabilities.

Next, we consider the ‘bracketing condition’ concerning the direct characteristics of the respective criteria, which are given by their likelihoods. Assuming – without loss of generality – that $m = +1$ ($m = -1$) is more likely for $w = B$ ($w = H$), the likelihoods of criterion i and i' fulfill

$$(A.10) \quad q_i^B > q_i^H \quad \text{and} \quad q_{i'}^B > q_{i'}^H.$$

Now, the bracketing condition reveals that the superior criterion i is characterized by a higher discriminatory power of the possible signal outcomes, which means that the signal outcomes can more probably be assigned to the true state of the world. Thus,

$$(A.11) \quad q_i^B > q_i^H \quad \text{and} \quad q_i^H < q_i^B.$$

The possible cases (i) and (ii) are described as follows:

$$\begin{aligned} (i) \quad EL_{i'}(h_j, d^0) &= q_{i'} \cdot L_H(1 - \hat{p}_{i'}(+1)) + (1 - q_{i'}) \cdot L_B \hat{p}_{i'}(-1) \\ &\stackrel{(1)}{=} L_H(1 - \hat{p}(h_j)) q_{i'}^H + L_B \hat{p}(h_j)(1 - q_{i'}^B) \\ &\stackrel{(B.11)}{>} L_H(1 - \hat{p}(h_j)) q_i^H + L_B \hat{p}(h_j)(1 - q_i^B) \\ &\stackrel{(1)}{=} q_i \cdot L_H(1 - \hat{p}_i(+1)) + (1 - q_i) \cdot L_B \hat{p}_i(-1) = EL_i(h_j, d^0); \end{aligned}$$

$$\begin{aligned} (ii) \quad EL_{i'}(h_j, d^0) &= q_{i'} \cdot L_B \hat{p}_{i'}(+1) + (1 - q_{i'}) \cdot L_B \hat{p}_{i'}(-1) \\ &\stackrel{(B.2)}{>} q_{i'} \cdot L_H(1 - \hat{p}_{i'}(+1)) + (1 - q_{i'}) L_B \hat{p}_{i'}(-1) \\ &\stackrel{(1)}{=} L_H(1 - \hat{p}(h_j)) q_{i'}^H + L_B \hat{p}(h_j)(1 - q_{i'}^B) \\ &\stackrel{(B.11)}{>} L_H(1 - \hat{p}(h_j)) q_i^H + L_B \hat{p}(h_j)(1 - q_i^B) \\ &\stackrel{(1)}{=} q_i \cdot L_H(1 - \hat{p}_i(+1)) + (1 - q_i) \cdot L_B \hat{p}_i(-1) = EL_i(h_j, d^0) \end{aligned}$$

$$\begin{aligned} \text{or} \quad EL_{i'}(h_j, d^0) &= q_{i'} \cdot L_H(1 - \hat{p}_{i'}(+1)) + (1 - q_{i'}) \cdot L_H(1 - \hat{p}_{i'}(-1)) \\ &\stackrel{(B.3)}{>} q_{i'} \cdot L_H(1 - \hat{p}_{i'}(+1)) + (1 - q_{i'}) \cdot L_B \hat{p}_{i'}(-1) \\ &\stackrel{(1)}{=} L_H(1 - \hat{p}(h_j)) q_{i'}^H + L_B \hat{p}(h_j)(1 - q_{i'}^B) \\ &\stackrel{(B.11)}{>} L_H(1 - \hat{p}(h_j)) q_i^H + L_B \hat{p}(h_j)(1 - q_i^B) \\ &\stackrel{(1)}{=} q_i \cdot L_H(1 - \hat{p}_i(+1)) + (1 - q_i) \cdot L_B \hat{p}_i(-1) = EL_i(h_j, d^0). \end{aligned}$$

Thus, the lower expected total losses of the superior criterion i are sufficiently derived – in each case – from condition (A.11).

References

- Athey, S., Levin, J. (2000): The Value of Information in Monotone Decision Problems, Working Paper at https://papers.ssrn.com/paper.taf?abstract_id=140591.
- Beckner, C. F., Salop, S. C. (1999): Decision Theory and Antitrust Rules, *Antitrust Law Journal* 67, 41-76.
- Ben-Bassat, M. (1978): Myopic Policies in Sequential Classification, *IEEE Transactions on Computers* C-27 (2), 170-174.
- Berger, J. O. (1993): *Statistical Decision Theory and Bayesian Analysis*, 2. ed. (3. corrected print.), New York: Springer.
- Bertocchi, G., Spagat, M. (1998): Growth under uncertainty with experimentation, *Journal of Economic Dynamics and Control* 23 (2), 209-231.
- Bielinska-Kwapisz, A. (2003): Sufficiency in Blackwell's theorem, *Mathematical Social Sciences* 46 (1), 21-25.
- Blackwell, D. (1951): Comparison of Experiments, in: Neyman, J. (ed.): *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, 93-102.
- Blackwell, D. (1953): Equivalent Comparison of Experiments, *The Annals of Mathematical Statistics* 24 (2), 265-272.
- Blackwell, D. and Girshick, M. A. (1966): *Theory of Games and Statistical Decisions*, New York: Wiley.
- Bradt, R. N., Karlin, S. (1956): On the Design and Comparison of Certain Dichotomous Experiments, *The Annals of Mathematical Statistics* 27 (2), 390-409.
- Crémer, J. (1982): A Simple Proof of Blackwell's "Comparison of Experiments" Theorem, *Journal of Economic Theory* 27 (2), 439-443.
- Cyert, R. M., DeGroot, M. H. (1987): *Bayesian Analysis and Uncertainty in Economic Theory*, Totowa, New Jersey: Rowman & Littlefield.
- Datta, M., Mirman, L. J., Schlee, E. E. (2000): Learning with noiseless information and payoff-relevant signals, *Economic Theory* 16 (1), 63-75.
- Datta, M., Mirman, L. J., Schlee, E. E. (2002): Optimal Experimentation in Signal-Dependent Decision Problems, *International Economic Review* 43 (2), 577-607.
- Davis, M. L. (1994): The Value of Truth and the Optimal Standard of Proof in Legal Disputes, *Journal of Law, Economics, and Organization* 10 (2), 343-359.
- DeGroot, M. H. (1961): Uncertainty, Information and Sequential Experiments, *The Annals of Mathematical Statistics* 33 (2), 404-419.
- DeGroot, M. H. (1970): *Optimal statistical decisions*, New York: McGraw-Hill.
- Drèze, J. H. (1960): Le paradoxe de l'information, *Economie Appliquée* 13 (1), 71-80.
- Drukarczyk, J. (1974): Zum Problem der Bestimmung des Wertes von Informationen, *Zeitschrift für Betriebswirtschaft* 44 (1), 1-18.
- Ferschl, F. (1982): Entscheidungstheoretische Grundlagen der Informationsökonomik, in: Streißler, E. (ed.): *Information in der Wirtschaft: Verhandlungen auf der Arbeitstagung des Vereins für Socialpolitik, Gesellschaft für Wirtschafts- und Sozialwissenschaften in Graz 1981*, Schriften des Vereins für Socialpolitik, Gesellschaft für Wirtschafts- und Sozialwissenschaften, Band 126, 35-78.

- Firchau, V. (1980): Bounds for the Expected Value of Information, *OR Spektrum* 2 (2), 65-73.
- Gorry, G. A. (1967): A system for computer-aided diagnosis, Project MAC, Technical Report MAC-TR-44 (Thesis), Cambridge, Massachusetts: MIT.
- Gorry, G. A. and Barnett, G. O. (1968): Experience with a Model of Sequential Diagnosis, *Computers and Biomedical Research* 1 (5), 490-507.
- Gould, J. P. (1974): Risk, Stochastic Preference, and the Value of Information, *Journal of Economic Theory* 8 (1), 64-84.
- Hadley, G. (1967): *Introduction to Probability and Statistical Decision Theory*, San Francisco: Holden-Day.
- Hamburg, M. (1970): *Statistical Analysis for Decision Making*, New York: Harcourt, Brace & World.
- Heckerman, D., Breese, J., Horvitz, E. (1989): The compilation of decision models, *Proceedings of Fifth Workshop on Uncertainty in Artificial Intelligence*, Windsor (Ontario), Association for Uncertainty in Artificial Intelligence, Mountain View (California), 162-173.
- Heckerman, D. E., Horvitz, E. J., Nathwani, B. N. (1992): Toward Normative Expert Systems: Part I, The Pathfinder Project, *Methods of Information in Medicine* 31 (2), 90-105.
- Heckerman, D. E., Nathwani, B. N. (1992): Toward Normative Expert Systems: Part II, Probability-Based Representations for Efficient Knowledge Acquisition and Inference, *Methods of Information in Medicine* 31 (2), 106-116.
- Heckerman, D., Horvitz, E., Middleton, B. (1993): An Approximate Nonmyopic Computation for Value of Information, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (3), 292-298.
- Howard, R. A. (1966): Information Value Theory, *IEEE Transactions on Systems Science and Cybernetics* 2 (1), 22-26.
- Hirshleifer, J., Riley, J. G. (1979): The analytics of Uncertainty and Information – An Expository Survey, *Journal of Economic Literature* 17 (4), 1375-1421.
- Hirshleifer, J., Riley, J. G. (1992): *The analytics of uncertainty and information*, Cambridge: Cambridge University Press.
- Kalagnanam, J., Henrion, M. (1990): A Comparison of Decision Analysis and Expert Rules for Sequential Diagnosis, *Uncertainty in Artificial Intelligence* 4, 271-281.
- Kerber, W., Kretschmer, J.-P., von Wangenheim, G. (2008): Optimal Sequential Investigation Rules in Competition Law, *Discussion Paper*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1266267
- Laux, H. (1993): *Entscheidungstheorie, Band 2: Erweiterung und Vertiefung*, 3. Auflage, Berlin: Springer.
- Lewis, C., Sheehan, K. (1990): Using Bayesian Decision Theory to Design a Computerized Mastery Test, *Applied Psychological Measurement* 14 (4), 367-386.
- Mag, W. (1977): *Entscheidung und Information*, München: Vahlen.
- Mannino, M. V., Mookerjee, V. S. (1999): Optimizing Expert Systems: Heuristics for Efficiently Generating Low-Cost Information Acquisition Strategies, *Journal on Computing* 11 (3), 278-291.

- Marschak, J. (1964): Problems in Information Economics, in: Bonini, C. P., Jaedicke, R. K., Wagner, H. M. (eds.): *Management controls: New directions in basic research*, New York: McGraw-Hill, 38-74.
- Marschak, J., Miyasawa, K. (1968): Economic Comparability of Information Systems, *International Economic Review* 9 (2), 137-174.
- Mookerjee, V. S., Dos Santos, B. L. (1993): Inductive Expert System Design: Maximizing System Value, *Information Systems Research* 4 (2), 111-140.
- Mookerjee, V. S., Mannino, M. V. (1997): Sequential Decision Models for Expert System Optimization, *IEEE Transactions on Knowledge and Data Engineering* 9 (5), 675-687.
- Moore, J. C., Whinston, A. B. (1986): A Model of Decision-Making with Sequential Information-Acquisition (Part 1), *Decision Support Systems* 2 (4), 285-307.
- Moore, J. C., Whinston, A. B. (1987): A Model of Decision-Making with Sequential Information-Acquisition (Part 2), *Decision Support Systems* 3 (1), 47-72.
- Morrison, C. T., Cohen, P. R. (2005): Noisy Information Value in Utility-based Decision Making, *Proceedings of the First International Workshop on Utility-based Data Mining*, 34-38.
- Motta, M. (2004): *Competition Policy: Theory and Practice*, Cambridge: Cambridge University Press.
- Mussi, S. (2002): Sequential decision-theoretic models and expert systems, *Expert Systems* 19 (2), 99-108.
- Ponssard, J.-P. (1975): A Note on Information Value Theory for Experiments Defined in Extensive Form, *Management Science* 22 (4), 449-454.
- Reid, M. C., Schoen, R. T., Evans, J., Rosenberg, J. C., Horwitz, R. I. (1998): The Consequences of Overdiagnosis and Overtreatment of Lyme Disease: An Observational Study, *Annals of Internal Medicine* 128 (5), 354-362.
- Sulganik, E., Zilcha, I. (1997): The value of information: The case of signal-dependent opportunity sets, *Journal of Economic Dynamics and Control* 21 (10), 1615-1625.
- Tseng, C.-C., Gmytrasiewicz, P. J. (1999): Time Sensitive Sequential Myopic Information Gathering, *IEEE Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1-7.
- Vanberg, V. (1993): Rational Choice, Rule-Following and Institutions, in: Gustafsson, B., Knudsen, C., Mäki, U. (eds.): *Rationality, Institutions and Economic Methodology*, London: Routledge Chapman & Hall, 171-200.
- Vos, H. J. (1999): Applications of Bayesian Decision Theory to Sequential Mastery Testing, *Journal of Educational and Behavioral Statistics* 24 (3), 271-292.
- Vos, H. J. (2007): A Bayesian sequential procedure for determining the optimal number of interrogatory examples for concept-learning, *Computers in Human Behavior* 23 (1), 609-627.
- Warner, H. R., Veasey, L. G. (1992): A Mathematical Approach to Medical Diagnosis: Application to Congenital Heart Disease, *M. D. Computing* 9 (1), 43-50.