
**CAUSE AND EFFECT IN VICTIM SENSITIVITY:
ANALYSES OF ASSOCIATED SOCIAL-COGNITIVE PROCESSES**

Dissertation

„kumulativ“

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Psychologie der Philipps-Universität Marburg

Vorgelegt von

Merle Buchholz

aus Göttingen

Marburg, 2021

Vom Fachbereich Psychologie der Philipps-Universität Marburg (Hochschulkennziffer 1180)
als Dissertation angenommen am 14.03.2021

Erstgutachter: Prof. Dr. Mario Gollwitzer, Ludwig-Maximilians-Universität München
Zweitgutachter: Prof. Dr. Erik Malte Müller, Philipps-Universität Marburg

Tag der Disputation: 22.03.2021

ACKNOWLEDGMENTS

Many people have helped and supported me during my PhD and I would like to acknowledge them all! First and foremost, I would like to thank Mario Gollwitzer for supervising me over the past three years. I am very grateful that you always made time for me when I had a question or needed advice. And thank you for all the great times I had in Munich and at the Kleinwalsertal retreats, I learned so much during those visits!

I would also like to thank all my colleagues in Marburg, Munich, and in the RTG 2271, especially Gunnar, Daniel, Erik, Sarah T.-M., Sofia, Zoe, Jone, and the one and only Po-Doc, Chris P. Thank you so much for all the help and support over the years!

A big thank you also to my Quod people, for all the great (and long) nights in “our” beer garden, for the Harry Potter tests, the fun, our weird conversations, and the virtual game nights during lockdown (and no, I’m not the Impostor)! I love you all! You guys have made this stressful time so much easier for me.

I am also exceptionally indebted to my family for their endless support. Thank you, Mama, Papa, Jonas, Leah, Omas and Opas, for all your love and encouragement and for always having my back. Without you, this thesis would not have been possible.

And last but not least, Evgenij. Thank you for everything. I know I will always be your little scientist.

CONTENTS

ABSTRACT.....	1
ZUSAMMENFASSUNG	2
ARTICLES	4
1 INTRODUCTION AND THEORETICAL BACKGROUND.....	5
1.1 Justice Sensitivity.....	6
1.2 The Sensitivity to Mean Intentions Model.....	10
1.3 Consequences of Victim Sensitivity	15
1.4 The Present Research.....	17
2 SUMMARIES OF THE MANUSCRIPTS.....	19
2.1 Victim Sensitivity in Patients with Borderline Personality Disorder.....	19
2.2 Cognitive and Motivational Processes Underlying Victim Sensitivity.....	21
2.3 Victim Sensitivity and Violations of (Un)Trustworthiness Expectations	27
3 DISCUSSION.....	34
3.1 Integration of Findings and Directions for Future Research.....	34
3.2 Methodological Limitations	40
3.3 Conclusion	42
REFERENCES	43
APPENDIX.....	61
Manuscript A	62
Manuscript B.....	91
Manuscript C.....	140
RELATIVE CONTRIBUTIONS TO THE PUBLICATIONS.....	184
CURRICULUM VITAE.....	185
EIGENSTÄNDIGKEITSERKLÄRUNG	188

ABSTRACT

Victim sensitivity (VS) – as one facet of justice sensitivity – is conceptualized as a personality disposition reflecting a combination of a strong need to trust others together with a latent expectation that other people are malevolent and untrustworthy. Previous research has linked victim sensitivity to a wide range of emotional and conduct problems that affect the respective individual as well as its social environment. However, there is still a lack of knowledge about how exactly VS shapes information processing and behavior.

This thesis adopts a social-cognitive perspective on victim sensitivity in order to gain a more complete understanding of the psychological processes that drive victim-sensitive people's reactions. To analyze underlying causes and effects, four studies were conducted that used timely and highly state-of-the-art procedures (e.g., eye tracking or virtual reality technology). Two important conclusions can be drawn from this research: first, the results confirm once again that victim sensitivity shapes emotions, cognitions, and behavior in a dysfunctional way. More specifically, the results suggested that VS is associated with hostile information processing (Study II) and contributes to even pathological forms of interpersonal dysfunction (Study I). Second, however, the present research also illustrates that it is possible to alleviate any adverse consequences of victim sensitivity. Study III, for example, found that individuals high in VS are able to overcome their habitual distrust if a sense of control is re-established. In addition, Study IV demonstrated that persons high in VS allocate preferential attention toward information violating their negative social expectations, which might reduce these expectations in the long run.

Taken together, the research described in this thesis extends our knowledge about social-cognitive and motivational processes underlying victim sensitivity and therefore has important implications for research on justice-related personality dispositions.

ZUSAMMENFASSUNG

Opfersensibilität (OS) – als eine Facette der Gerechtigkeitssensibilität – wird als Persönlichkeitsdisposition konzeptualisiert, welche eine Kombination aus einem starken Vertrauensmotiv sowie einer latenten Erwartung, dass andere Menschen böswillig und vertrauensunwürdig sind, widerspiegelt. Frühere Forschung hat Opfersensibilität mit einer Vielzahl von emotionalen und Verhaltensproblemen in Verbindung gebracht, die sowohl Auswirkungen auf das jeweilige Individuum als auch auf sein soziales Umfeld haben. Bislang mangelt es allerdings noch an Wissen darüber, wie genau OS die Informationsverarbeitung und das Verhalten prägt.

In dieser Dissertation wird eine sozial-kognitive Perspektive auf Opfersensibilität eingenommen, um ein umfassenderes Verständnis der psychologischen Prozesse zu erlangen, die die Reaktionen opfersensibler Menschen formen. Um die zugrundeliegenden Ursachen und Wirkungen zu analysieren, wurden vier Studien durchgeführt, die hochmoderne Verfahren (z.B. Eye-Tracking oder Virtual-Reality-Technologie) verwendeten. Zwei wichtige Schlussfolgerungen können aus dieser Forschung gezogen werden: Erstens bestätigen die Ergebnisse, dass Opfersensibilität Emotionen, Kognitionen und Verhalten in einer dysfunktionalen Art und Weise beeinflusst. Insbesondere legen die Ergebnisse nahe, dass OS mit feindseliger Informationsverarbeitung (Studie II) assoziiert ist und sogar zu pathologischen Formen zwischenmenschlicher Dysfunktion beiträgt (Studie I). Zweitens zeigt die vorliegende Untersuchung allerdings auch, dass es möglich ist, die negativen Folgen von Opfersensibilität abzuschwächen. So wurde in Studie III festgestellt, dass Personen mit hoher OS in der Lage sind, ihr habituelles Misstrauen zu überwinden, wenn ein Gefühl der Kontrolle wiederhergestellt wird. Darüber hinaus konnte in Studie IV gezeigt werden, dass Personen mit

hoher OS ihre Aufmerksamkeit bevorzugt auf Informationen richten, die ihre negativen sozialen Erwartungen verletzen, was diese Erwartungen langfristig reduzieren könnte.

Zusammengenommen erweitert die in dieser Arbeit beschriebene Forschung unser Wissen über sozial-kognitive und motivationale Prozesse, die der Opfersensibilität zugrunde liegen und hat daher wichtige Implikationen für die Forschung zu gerechtigkeitsbezogenen Persönlichkeitsdispositionen.

ARTICLES

This cumulative thesis is based on three manuscripts submitted for publication in peer-reviewed journals that will be discussed in the order in which they are listed below. A summary of the articles' theoretical frameworks, methods, and results is given in the main text of this thesis; for details the reader is referred to the complete manuscripts as appended to this work.

Buchholz, M., Neukel, C., Steinmann, S., Lis, S., & Gollwitzer, M. (submitted). Measuring justice sensitivity in patients with borderline personality disorder: Do we really measure the same? Manuscript submitted for publication in *Assessment*.

Buchholz, M., Magraw-Mickelson, Z., Stolz, C., Süßenbach, P., & Gollwitzer, M. (submitted). Taking back control: Findings on the cognitive, behavioral, and motivational consequences of victim sensitivity. Manuscript submitted for publication in *Social Justice Research*.

Buchholz, M., Bergmann, N., Schubö, A., & Gollwitzer, M. (submitted). Victim sensitivity predicts attention allocation towards violations of untrustworthiness expectancies. Manuscript submitted for publication in *Social Justice Research*.

Funding: The research for this Doctoral Dissertation was conducted in the context of the DFG-funded Research Training Group 2271 “Expectation Maintenance vs. Change in the Context of Expectation Violations: Connecting Different Approaches” (project number 290878970 – GRK2271 – Project 4).

1 INTRODUCTION AND THEORETICAL BACKGROUND

Justice is a central topic in human social life. In organizational contexts, for example, people expect equal pay for equal work; otherwise, they feel treated unfairly. Similarly, if one learns that another applicant is favored in the selection process only because he or she is an acquaintance of the company's manager, this will be perceived as undeserved. In courts and parliaments, outcomes and procedures for decision-making are rated as unjust if not every party has the right to voice an opinion. Finally, justice issues are at the core of everyday conflicts in families and romantic relationships, and even lie at the heart of wars, revolutions, and political protests.

All of these examples illustrate that humans care deeply about justice. In fact, justice research in psychology and related disciplines has identified a *justice motive* that is considered to be anthropologically universal (Dalbert, 2001; Lerner, 1977; Montada, 2007)¹. Thus, people have a desire for justice to prevail in the world, which implies that everyone should deserve what they get, and get what they are entitled to (Lerner, 1980). In addition, people want to perceive themselves as doing the right thing, i.e. to being moral, fair, and benign human beings (Gollwitzer & van Prooijen, 2016; Jordan et al., 2011).

To explain why and when people strive for justice for themselves and others, and under which circumstances they feel treated (un)fairly, influential theories like equity theory (Adams, 1963, 1965), relative deprivation theory (Crosby, 1976), and justice motive theory (Lerner, 1977, 1980) have been developed in the last 50 to 60 years. However, situational factors alone cannot fully explain variance in justice-related cognitions, emotions, and behaviors: individuals

¹ In psychology, motives are for example defined as “a disposition to be concerned with and to strive for a certain class of incentives or goals” (Emmons, 1989, p.32) or as “psychological processes that propel people’s thinking, feeling, and behaving” (Fiske, 2018, p.12).

greatly differ in their concern for justice, even if placed in the same objective situation (Baumert & Schmitt, 2016; Gollwitzer & van Prooijen, 2016; Huseman et al., 1987; Mohiyeddini & Schmitt, 1997; Schmitt, 1996; Schmitt & Mohiyeddini, 1996; van den Bos et al., 2003). For this reason, contemporary social justice research has also begun to consider stable inter-individual differences in the desire for justice (e.g., Baumert et al., 2013; Dalbert, 2001; Montada et al., 1986; Schmitt et al., 1995).

1.1 Justice Sensitivity

Justice is a fundamental human motivation but similarly to other motives, its strength varies between individuals. One construct that is supposed to reflect inter-individual differences in the strength of the justice motive is *justice sensitivity* (JS; Schmitt et al., 2005; Schmitt et al., 2010)². As proposed by Schmitt, Neumann, and Montada (1995), this personality disposition captures how readily people perceive and how strongly they react to injustice in general. More specifically, individuals who are particularly justice-sensitive are assumed to have a lower perceptual threshold for contextual cues related to (in)justice and should thus experience (and also remember) injustice more frequently than insensitive persons (Schmitt, 1996; Schmitt et al., 1995). In addition, Schmitt et al. (1995) posit that people high (vs. low) in justice sensitivity show stronger emotional reactions following the detection of injustice – in the case of unfair treatment for the self, this emotion is typically anger (Mikula et al., 1998). Because strong negative emotions (like anger) tend to result in intrusive thoughts about the respective event (Rimé et al., 1992), justice-sensitive persons should also ruminate longer about

² Another construct that has been proposed as an indicator for the justice motive is the *belief in a just world* (Rubin & Peplau, 1973, 1975). However, although both constructs are conceptualized to capture the degree of people's justice concerns, they are only weakly correlated (e.g., Schmitt et al., 2005). Baumert et al. (2013) therefore argue that both constructs reflect different facets of the justice motive: while justice sensitivity may be seen as indicative of unconditional justice strivings in the sense of a moral motivation, the belief in a just world rather represents a conditional need for a just (and therefore controllable) world.

unfair incidents than justice-insensitive persons. Finally, individuals high in justice sensitivity are assumed to be especially motivated to redress injustice, for example by means of retributive and punitive acts (Baumert & Schmitt, 2016; Schmitt et al., 1995).

Based on these four components or indicators that are assumed to constitute the JS construct (i.e., frequency of perceived injustice, affective reactivity, extent of rumination, punitivity), Schmitt et al. (1995) initially developed a self-report questionnaire to measure justice sensitivity from a victim's perspective³. As expected, justice sensitivity as assessed by this scale (or, *sensitivity to befallen injustice*, as the construct was called at that time) was found to predict emotional, cognitive, and behavioral reactions to subjective injustice above and beyond other relevant constructs. For example, Schmitt and Mohiyeddini (1996) showed that students high in sensitivity to own unjust disadvantages judged a lottery employed to allocate scarce teaching resources as more unjust, and that they approved more strongly of activities aimed at replacing the procedure. Similarly, Mohiyeddini and Schmitt (1997) found justice sensitivity from a victim's perspective to be a better predictor of unfairness judgments, resentment and protests against an unfair procedure than related constructs like trait anger or self-assertiveness. Moreover, Schmitt and Dörfel (1999) demonstrated that (victim) justice sensitivity amplified negative effects of procedural injustice on psychosomatic well-being.

All of these findings support the notion that people differ systematically in how sensitive they are to being unjustly disadvantaged, and that these inter-individual differences can account for individuals' reactions to perceived injustice. It should be noted, though, that an unjust incident typically does not only involve the person suffering from the unfairness. Most of the time, other people are directly responsible for the injustice or passively profit from it as beneficiaries. In addition, one or more uninvolved bystanders may witness the unfair incident.

³ In contrast to the first version by Schmitt et al. (1995), recent versions of the justice sensitivity inventory (Schmitt et al., 2005; Schmitt et al., 2010) have excluded items referring to the indicators *frequency of perceived injustice* and *punitivity* for efficiency reasons. Nevertheless, a low activation threshold and a desire for retribution are still assumed to be relevant components of the JS construct (see also Baumert & Schmitt, 2016).

Thus, people can adopt different roles in justice-relevant situations (Mikula, 2002). Because affective, cognitive, and behavioral reactions to unfair treatment differ greatly depending on the perspective from which an injustice is experienced (e.g., Weiss et al., 1999), justice sensitivity has recently been decomposed into the four dimensions victim, observer, beneficiary, and perpetrator sensitivity (Schmitt et al., 2005; Schmitt et al., 2010). In other words, people may not only differ in how sensitively they react toward own unjust disadvantages, but also in how readily they perceive injustice that is inflicted upon other people, either by themselves or by third parties.

The newest and most prominent version of the self-report scales for justice sensitivity (Schmitt et al., 2010) has been shown to assess the four dimensions reliably and validly⁴. Importantly, the JS facets have been shown to be highly stable and sufficiently distinct from other personality dispositions, such as (un)just world beliefs or the Big Five personality factors (Schmitt et al., 2005; Schmitt et al., 2010). In addition, the JS facets are positively inter-correlated, supporting the notion that they capture a general concern for justice. At the same time, the pattern of correlations reflects qualitative differences between perspectives as well: while some facets are more closely related, others share less similarities. For example, beneficiary and perpetrator sensitivity are most highly correlated, which corresponds to the fact that these perspectives have several elements in common (Baumert & Schmitt, 2016; Schmitt et al., 2010). Passive beneficiaries as well as active perpetrators both profit from an unjust incident at the expense of others, which can result in feelings of guilt or shame and in strivings to punish the self and/or to compensate the victim (Baumeister et al., 1994; Klass, 1978; Montada et al., 1986; Nelissen & Zeelenberg, 2009; Weiss et al., 1999). In contrast, bystanders are not directly involved but typically experience moral outrage on behalf of the victim (Lotz

⁴ Confirmatory factor analyses have corroborated the assumed four-factor structure (Schmitt et al., 2010). The new scales measure each facet with ten items each that are parallel in wording but differ with regard to the perspective from which an injustice is perceived.

et al., 2011; Montada, 1993; Montada & Schneider, 1989). Accordingly, correlations with observer sensitivity were found to be lower. However, the smallest – although significant – correlations exist between the beneficiary and perpetrator facets and victim sensitivity, as these roles imply directly opposing outcomes (benefitting vs. suffering from an injustice), and therefore a negative interdependency (Schmitt et al., 2010).

Research investigating the relation between justice sensitivity and other constructs as well as studies exploring the effects of JS on justice-related emotions, cognitions, and behavior have further highlighted distinctions in the psychological meaning of the JS perspectives. More specifically, it has been shown that observer, beneficiary, and perpetrator sensitivity are associated with prosocial behavior and other-related concerns, while victim sensitivity is rather related to antisocial tendencies and self-related concerns. For example, victim sensitivity was found to be positively correlated with paranoia, hostility, vengeance, jealousy, and neuroticism, whereas the other three facets show positive correlations with constructs such as empathy, modesty, tender-mindedness, social responsibility, norm compliance, and solidarity with the disadvantaged (Gollwitzer et al., 2005; Schmitt et al., 2005; Schmitt et al., 2010)⁵.

Victim sensitivity has also been associated with a higher willingness to transgress in tempting situations, i.e. with less adherence to fairness norms. In a study by Faccenda and colleagues (2009), for instance, soccer players high in victim sensitivity were more inclined to adopt unsportsmanlike, transgressive behaviors after imagining an unfair refereeing decision. Participants high in observer and beneficiary sensitivity, however, showed high moral intentions and refrained from using deviant means. Similarly, research in the context of economic games (Fetchenhauer & Huang, 2004; Lotz et al., 2013) has demonstrated that participants high (vs. low) in victim sensitivity tend to behave selfishly and that they propose

⁵ In some of the earlier work (e.g., Fetchenhauer & Huang, 2004; Schmitt et al., 2005) beneficiary sensitivity was called “perpetrator sensitivity”. The facet was officially renamed to beneficiary sensitivity when Schmitt et al. (2010) introduced the differentiation between actively committing vs. passively benefitting from an unfairness.

unequal splits of financial endowments between themselves and other persons. In contrast, subjects high in observer, beneficiary, and perpetrator sensitivity were not only found to be less willing to accept unequal offers for themselves or others, but also to remain altruistic even if tempted to do otherwise (e.g., when offering money to a recipient who does not know that his or her outcome depends on the decision of another person).

In conclusion, these and similar findings (see also Gerlach et al., 2012; Gollwitzer & Rothmund, 2011; Gollwitzer et al., 2009; Lavelle et al., 2018) suggest that observer, beneficiary, and perpetrator sensitivity capture an intrinsic, other-oriented concern for justice. Victim sensitivity shares common variance with these facets and hence reflects a desire for justice as well, at least to some extent. However, being sensitive to injustice from a victim's perspective also predicts selfish and uncooperative behavior and therefore seems to involve a mixture of egoistic and justice-related strivings (Schmitt et al., 2005). For this reason, it has been proposed that victim-sensitive persons care more about *justice for themselves* than about justice for others (Gollwitzer et al., 2005).

1.2 The Sensitivity to Mean Intentions Model

At first glance, it seems puzzling that a personality disposition that has been conceptualized as an indicator for the justice motive results in uncooperative and even immoral behavior in various social contexts. So how can this contradiction be resolved? An answer to this question is provided by the Sensitivity to Mean Intentions (SeMI) model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013), which is graphically depicted in Figure 1. According to this theoretical framework, victim-sensitive individuals tend to behave selfishly, or rather, defensively, because they are afraid that they will otherwise become the victim of other people's mean intentions.

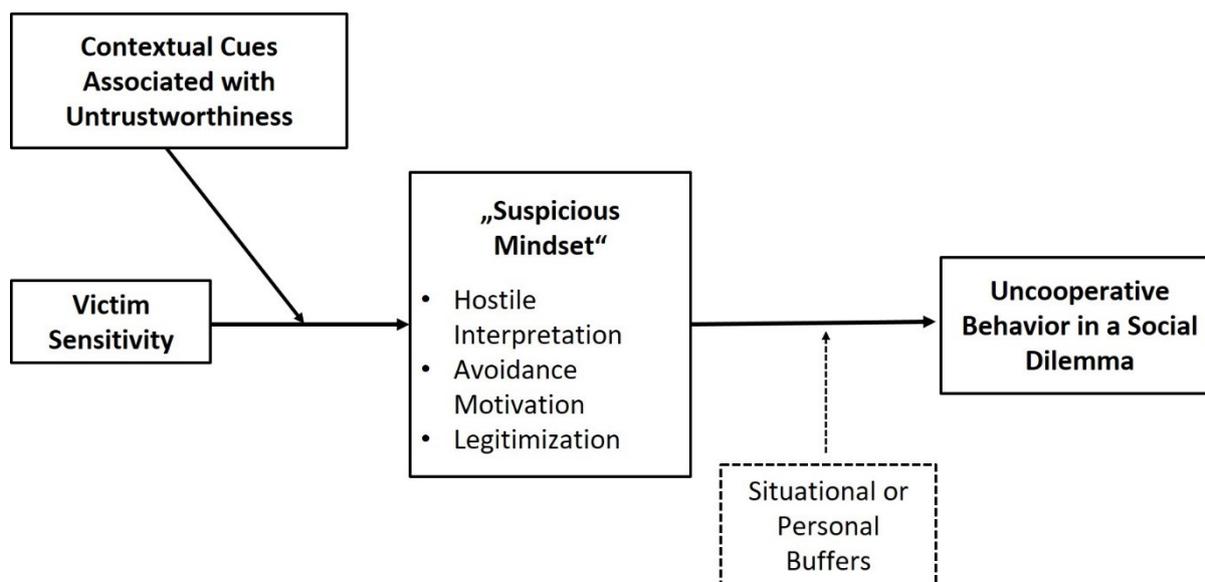


Figure 1. The Sensitivity to Mean Intentions Model as described in Gollwitzer et al., 2013. Reprinted by permission.

In more detail, the model posits that people high in victim sensitivity (VS) are particularly sensitive toward the possibility of being exploited by others in socially uncertain situations. This sensitivity is assumed to be rooted in a specific dissonance between (1) a strong need to trust others (e.g., Fiske, 2018) together with (2) a generalized expectation that other people are malevolent and untrustworthy. Thus, victim-sensitive persons expect that other people will take advantage of them, and, at the same time, they are particularly averse to this expectation because it violates their need to trust. Stated differently: individuals high in VS are assumed to harbor a latent *fear of exploitation* (also called *sugrophobia*, c.f. Vohs et al., 2007) and, for this reason, should be strongly motivated to avoid being victimized by others⁶.

Unfortunately, social situations offer many opportunities for exploitation. Therefore, people are generally well advised to not trust others blindly, but to look for information that

⁶ This conceptualization distinguishes victim sensitivity from related concepts such as trait trust. *Interpersonal* or *general trust* is often defined as a “generalized expectancy that the oral or written statements of other people can be relied upon” (Rotter, 1967, p. 653) or as an “expectation of goodwill and benign intent” (Yamagishi & Yamagishi, 1994, p. 131). Thus, trait trust also captures generalized expectations regarding other people’s (un)trustworthiness, but it does not entail a motivational aspect (e.g., no *need* to trust). As a consequence, VS and general trust are only moderately correlated (Baumert, Beierlein et al., 2014; Schmitt et al., 1995; Schmitt et al., 2005).

can help to decide whether an interaction partner is actually trustworthy (Gollwitzer et al., 2013). Accordingly, if contextual cues suggest that another person has ulterior motives, then it would probably be safer not to take any chances, i.e., to withdraw one's trust and cooperation. Such indications for untrustworthiness may for example include hostile facial expressions (e.g., Oosterhof & Todorov, 2009; Todorov, 2008; Todorov et al., 2009; Zebrowitz & Montepare, 2008) or the explicit disclosure of mean intentions (e.g., Parks et al., 1996), but also more subtle behaviors like gaze aversion (e.g., Vrij et al., 2001).

Importantly, the SeMI model posits that victim-sensitive people – who strongly worry about being deceived by others – are especially sensitive towards cues suggesting that other people harbor malevolent intentions. Support for this notion has been found in a number of studies. For example, Gollwitzer et al. (2009) instructed their participants to play a social dilemma game⁷ after watching some sample rounds in which either none, some, or many players violated fairness standards and defected. In line with the SeMI model's predictions, victim-sensitive participants (but not victim-insensitive participants) reacted to even weak cues of untrustworthiness and invested less of their own endowment if they had been confronted with only a few violators before. In a similar way, Gollwitzer and colleagues (2012) demonstrated that people high in VS perceived faces with neutral and hostile (i.e., untrustworthy) expressions as more untrustworthy than people low in VS. Ratings for friendly (i.e., trustworthy) looking faces, however, did not differ between participants high and low in victim sensitivity. In other words, victim-sensitive participants responded more sensitively to untrustworthiness-related than to trustworthiness-related information, and they even tended to interpret very ambiguous cues (i.e., neutral expressions) as indicative of mean intentions.

⁷ A social dilemma is defined as a choice between a selfish and a cooperative alternative in a social context, i.e., as a conflict between self-interests and collective interests (e.g., Dawes, 1980; De Cremer et al., 2001; Gollwitzer & Rothmund, 2009). Importantly, such situations imply (1) uncertainty: no actor can be sure how the others are going to decide; and (2) interdependence between all actors: although every actor can maximize their payoff by behaving uncooperatively, all individuals are worse off if all behave egoistically. Stated differently: cooperation can be costly but pays off if others cooperate as well.

Together, these findings corroborate the SeMI model's assumption that victim sensitivity is associated with an *asymmetrical sensitivity* towards untrustworthiness cues, although it may be doubtful whether these clues always have prognostic validity for predicting exploitation (e.g., in the case of a neutral facial appearance)⁸.

When untrustworthiness cues are detected in a given situation, the model further predicts that victim-sensitive people adopt a so-called *suspicious mindset*, which then influences subsequent behavior. The SeMI model hence describes a moderated mediation: victim-sensitive people act uncooperatively if untrustworthiness cues are present in a social dilemma situation (moderator) because these cues trigger a state of suspiciousness (mediator), and, thus, reduce norm compliance. The suspicious mindset is assumed to consist of three components (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). As can be seen in Figure 1, the first component is an attributional bias toward malevolence: after encountering untrustworthiness cues in a social context, victim-sensitive persons are assumed to ascribe sinister motives to others, i.e., to interpret other people's behavior and intentions in a more hostile way. The second element is avoidance motivation. According to the SeMI model, individuals high in VS are inclined to make special efforts to avoid exploitation whenever they expect that an interaction partner harbors mean intentions. The third component is defined as a heightened accessibility to legitimizing cognitions. Put differently: in a state of suspiciousness, victim-sensitive people are assumed to justify their own selfish behavior by arguing that they would otherwise be unjustly disadvantaged themselves. These legitimizing, protective cognitions may thus serve to reduce feelings of guilt and self-blame that usually accompany the violation of moral principles (Bandura, 1991; Gollwitzer et al., 2013; Klass, 1978).

⁸ Notably, findings by Baumert et al. (2012) suggest that this sensitivity may be understood as an integrational rather than as an attentional asymmetry. In their study, VS was related to the preferential processing of both unfairness- and fairness-related information. Accordingly, victim-sensitive people do seem to perceive and attend to fairness cues as well, but to rely more strongly on unfairness cues when forming trustworthiness judgments. However, more research is needed in this context as Baumert and colleagues focused on justice- and injustice-related information in a broader sense and not on cues specifically related to (un)trustworthiness.

First evidence for the suspicious mindset hypothesis has been provided by a number of studies. For example, Gerlach et al. (2012) investigated the suspicious mindset in research on transgressions in close relationships and found that (1) participants high in VS were less willing to forgive the perpetrator and that (2) hostile interpretations and legitimizing cognitions partially mediated the relationship between victim sensitivity and unforgiving reactions. Following these findings, Agroskin et al. (2015) applied the Sensitivity to Mean Intentions model to the political domain. As expected, the results demonstrated that victim-sensitive participants were more inclined than victim-insensitive participants to oppose political reforms because they ascribed ulterior motives to the policy makers. This attributional bias, however, depended on the presence of untrustworthiness cues: without any evidence for untrustworthiness, no suspiciousness was evoked and thus, no effect of VS on resistance to reforms was found. Finally, Maltese and colleagues (2016) corroborated the suspicious mindset hypothesis in the context of economic games. More specifically, they showed that participants high (vs. low) in victim sensitivity were more likely to expect injustice after being confronted with unfairness cues, and that these hostile interpretations mediated the effect of VS on uncooperative decisions (for similar results, see Rothmund et al., 2011). Thus, past research has corroborated the notion that a suspicious mindset is easily activated in victim-sensitive people and that such a state of suspiciousness results in an attributional bias toward malevolence. Support for the other two components (i.e., avoidance motivation, legitimization), however, is scarce.

In conclusion, the SeMI model describes social-cognitive and motivational processes that can explain why and when victim sensitivity translates into uncooperative or even antisocial behavior. Testable predictions can be derived from this model, some of which have already been tested in previous research. Most importantly, the model assumes that in socially uncertain situations, individuals high in VS more easily become suspicious, which makes them

behave defensively in order to protect themselves from being exploited by others. However, it should be noted that the activation of a suspicious mindset does not *necessarily* result in the withdrawal of trust and cooperation. Certain buffers – such as situational constraints or high moral standards – may mitigate the effects of suspiciousness on behavior (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). In other words, victim-sensitive people may sometimes feel obligated to adhere to fairness norms even though they do not expect their counterparts to do the same.

1.3 Consequences of Victim Sensitivity

Victim-sensitive people are particularly sensitive toward untrustworthiness cues and behave pre-emptively selfishly if they anticipate being taken advantage of by others. In other words, being vigilant against victimization can help to detect malicious intentions and to avoid exploitation, and may thus be considered adaptive from an evolutionary perspective (Cosmides & Tooby, 1992, 2005). However, although a high fear of exploitation can be functional at times, there are negative side effects as well. For example, while decreasing the probability for “type-1-errors” (i.e., trusting people who should not be trusted), victim sensitivity simultaneously increases the probability of “type-2-errors” (i.e., not trusting others who are actually trustworthy) (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013; Schmitt et al., 2005). In fact, a study by Gollwitzer et al. (2012) demonstrated that individuals high in VS tend to systematically underestimate other people’s cooperativeness. This bias in social judgments, on the other hand, might crucially impair social interactions: if people high in VS behave uncooperatively because they expect others to be untrustworthy, then their interaction partners should be more hesitant to cooperate in return. Thus, by being overly suspicious, victim-sensitive individuals may help to create vicious cycles of mutual distrust and non-cooperation (Darley & Fazio, 1980; Gollwitzer et al., 2015). In line with this reasoning, victim

sensitivity has been shown to aggravate interpersonal conflict and to undermine conflict resolution (Baumert, Nazlic, et al., 2014; Baumert & Schmitt, 2016; Gerlach et al., 2012).

Importantly, VS not only has negative interpersonal consequences, but also adverse effects on the individual's mental health. For example, it has been argued that a very strong fear of exploitation may culminate in pathological forms of mistrust, like paranoid personality disorder (e.g., Carroll, 2009; Gollwitzer et al., 2013). In addition, there is consistent evidence that victim sensitivity represents a risk factor for conduct problems such as aggressiveness (Bondü & Elsner, 2015; Bondü & Krahe, 2015; Bondü & Richter, 2016; Lovas & Wolt, 2002), and for decreased psychological well-being and life satisfaction (Baumert, Beierlein et al., 2014; Schmitt & Dörfel, 1999; Schmitt et al., 1995). Finally, victim sensitivity was found to be higher in children with ADHD symptoms (Bondü & Esser, 2015) and in adults with elevated borderline personality disorder features (Lis et al., 2018), and has been linked to the stabilization of depressive symptoms (Bondü et al., 2017).

In summary, these findings demonstrate that victim sensitivity is related to a wide range of emotional and behavioral problems that affect the respective individual as well as its social environment. However, although past research has repeatedly identified victim sensitivity as a risk factor for reduced psychosocial functioning, much less is known about how these effects may be prevented. Further investigating the social-cognitive mechanisms underlying a heightened sensitivity to mean intentions is therefore of utmost importance: if we understand *how exactly* victim-sensitive people think, feel and act in social situations, and *why* they do so, we might find ways to mitigate the adverse consequences of victim sensitivity. The present research takes important first steps in this regard.

1.4 The Present Research

Although there has been a substantial amount of research on justice sensitivity in general, and on victim sensitivity in particular, there are still numerous empirical questions that await answering. Most importantly, previous studies have focused primarily on the behavioral consequences of being dispositionally victim-sensitive, while underlying social-cognitive and motivational processes, which are for example proposed by the SeMI model (e.g., the suspicious mindset), have received far less attention. In other words, there is still a lack of knowledge about *how exactly* VS shapes information processing and behavior. In addition, even though victim sensitivity has received considerable attention in the field of social psychology, less studies have focused on the role of VS in clinical and psychopathological contexts (but see for example Bondü & Esser, 2015; Bondü et al., 2017; Lis et al., 2018). It is therefore not yet fully clarified whether the victim sensitivity construct can contribute to the understanding of dysfunctional behavior patterns associated with mental disorders such as borderline personality disorder (BPD). The present research specifically addresses these research gaps.

This thesis consists of four studies presented in three manuscripts that were designed to tackle different research questions with regard to the behavioral, cognitive, and motivational consequences of victim sensitivity. In more detail, the aim of Study I was to follow up on results of Lis et al. (2018), who found preliminary evidence for the assumption that victim sensitivity may be a meaningful construct in the context of borderline personality disorder. To further examine the relationship between VS and BPD symptomatology, we compared latent VS scores between a clinical sample of BPD patients diagnosed by trained psychologists with standardized instruments and a sample of matched controls. To do so, we used the alignment method (Asparouhov & Muthén, 2014) and also tested for (approximate) measurement invariance.

Studies II and III were subsequently implemented to gain a more complete understanding of the psychological processes that drive victim-sensitive people's reactions. In other words, these two studies were designed to determine *how* (Study II) and *why* (Study III) victim sensitivity translates into socially dysfunctional behavior. More specifically, Study II looked for stronger and more direct support for the conceptualization of the suspicious mindset as described in the SeMI model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). To overcome the limitations of previous research, we simultaneously investigated the mediating effects of all three elements that are assumed to constitute this mindset (i.e., attributional bias, avoidance motivation, legitimization) in an ecologically realistic virtual environment. Study III complemented this research by investigating what *motivates* people high in victim sensitivity to behave selfishly in social situations, with the goal to find effective ways to change dysfunctional behavior patterns. More precisely, in Study III we implemented and compared two different approaches (a self-affirmation procedure and a control-affirmation procedure) that, according to theoretical considerations, seemed promising for mitigating the negative consequences of victim sensitivity.

Finally, Study IV went a step further than previous research on victim sensitivity and examined how people high in VS process *expectation-inconsistent* information regarding (un-)trustworthiness. More specifically, Study IV tested whether victim-sensitive individuals allocate preferential attention toward social information that violates trustworthiness expectations, or toward social information that violates untrustworthiness expectations. To do so, we used eye tracking; an empirical approach that is timely and highly state-of-the-art.

2 SUMMARIES OF THE MANUSCRIPTS

2.1 Victim Sensitivity in Patients with Borderline Personality Disorder

Buchholz, M., Neukel, C., Steinmann, S., Lis, S., & Gollwitzer, M. (submitted). Measuring justice sensitivity in patients with borderline personality disorder: Do we really measure the same? Manuscript submitted for publication in *Assessment*.

Feelings of injustice are highly aversive and have detrimental effects on a person's physical and mental health (Elovainio et al., 2002; Greenberg, 2006, 2010). People who are prone to experience unjust treatment frequently and intensely should thus be at high risk for psychiatric morbidity. In line with this reasoning, research in clinical psychology and psychopathology has linked a heightened sensitivity to injustice to borderline personality disorder (BPD) symptomatology (e.g., Lis et al., 2018). BPD is a severe mental disorder associated with interpersonal deficits and emotional problems, including intense anger reactions and aggressive behaviors, an unstable sense of self, impulsivity, and self-harm (Bateman & Krawitz, 2013; Lazarus et al., 2014; Mancke et al., 2015; Schmahl et al., 2014). Importantly, the majority of diagnosed BPD patients report past experiences of sexual and/or physical abuse, and thus, a history of victimization (Ball & Links, 2009; Battle et al., 2004; Herman et al., 1989; Ogata et al., 1990; Zanarini et al., 1997). Because of these past traumas, it has been argued that people with BPD react very sensitively (and in dysfunctional ways) to injustice inflicted upon themselves or others (Bateman & Krawitz, 2013). Stated differently: the interpersonal problems associated with BPD, such as angry and aggressive reactions, may result – at least in part – from heightened justice concerns and from a strong desire to redress

injustice no matter the costs. Especially victim sensitivity with its close link to anger and antisocial behavior may thus be a relevant factor contributing to the interpersonal dysfunction of BPD patients. In line with this notion, Lis and colleagues (2018) demonstrated that (1) participants with a clinically relevant degree of BPD features showed higher mean values on victim sensitivity (and observer sensitivity) than participants with less pronounced BPD features and that (2) VS partially mediated the relation between BPD features and aggression.

Study I

In Study I, we expanded these findings in two important ways. First, we compared BPD patients ($n_1=91$) with healthy controls ($n_2=455$) that matched the patient sample in terms of age, sex, and education. Thus, our study was the first to investigate victim sensitivity and the other JS facets in a clinical sample where BPD was diagnosed by trained psychologists according to DSM-IV criteria (American Psychiatric Association, 2000). Second, we used the new and promising alignment procedure (Asparouhov & Muthén, 2014)⁹ to test for (approximate) measurement invariance of the justice sensitivity scales (Schmitt et al., 2005; Schmitt et al., 2010). This was necessary because measurement invariance is a key requirement if (latent) means are to be meaningfully compared between groups, such as BPD patients and healthy controls (Brown, 2006; Davidov, 2010; Steenkamp & Baumgartner, 1998; van de Schoot et al., 2015). In other words, investigating victim sensitivity in the context of BPD first requires showing that the construct is measured comparably in both groups; otherwise, differences in mean scores could simply be due to different response styles, that is, to methodological artefacts (e.g., van de Schoot et al., 2012).

⁹ In contrast to more traditional approaches like multi-group confirmatory factor analysis, the alignment procedure allows to estimate trustworthy means and variances even if a slight degree of non-invariance remains (i.e., measurement parameters must be sufficiently similar, but not identical across groups). In addition, alignment greatly simplifies measurement invariance testing (Asparouhov & Muthén, 2014; Byrne & van de Vijver, 2017; Muthén & Asparouhov, 2014).

Our alignment results found the justice sensitivity scales to be approximately invariant across groups, which ensures that BPD patients and matched controls responded comparably to the items. Mean differences between groups can thus not be attributed to methodological artefacts. Even more importantly, we replicated the findings by Lis et al. (2018) and demonstrated increased levels of victim sensitivity in the clinical sample compared to the control sample¹⁰. These results, which reflect *true-score differences* in victim sensitivity, hence confirm heightened justice concerns as an important characteristic of BPD psychopathology. In other words, Study I further corroborates the notion that VS adds to the social and emotional problems of persons with BPD, and, as such, may be considered a risk factor for even pathological forms of interpersonal dysfunction. This again highlights the need for further research, particularly with regard to the social-cognitive processes that translate victim sensitivity into dysfunctional reactions and behavior.

2.2 Cognitive and Motivational Processes Underlying Victim Sensitivity

Buchholz, M., Magraw-Mickelson, Z., Stolz, C., Süssenbach, P., & Gollwitzer, M. (submitted). Taking back control: Findings on the cognitive, behavioral, and motivational consequences of victim sensitivity. Manuscript submitted for publication in *Social Justice Research*.

As previous research and the results of Study I suggest, victim sensitivity has detrimental effects on a person's social functioning. However, whereas a number of studies have looked at the adverse behavioral consequences of VS (e.g., Faccenda et al., 2009; Fetchenhauer & Huang, 2004; Gollwitzer et al., 2005; Gollwitzer & Rothmund, 2011; Lavelle et al., 2018), few have investigated the underlying cognitive and motivational mechanisms. To

¹⁰ BPD patients also showed higher levels of observer and beneficiary sensitivity than healthy controls, illustrating that these facets may have detrimental effects as well. Perpetrator sensitivity was excluded from all analyses because of numerous missing values on this facet among the healthy controls.

close this research gap, we conducted two studies (Study II and Study III) which are briefly described below.

Study II

According to the Sensitivity to Mean Intentions model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013), victim-sensitive people are characterized by a strong fear of exploitation, which can explain why they behave uncooperatively and even antisocially in social dilemma situations. More specifically, victim-sensitive individuals are assumed to adopt a suspicious mindset whenever they have reason to believe that an interaction partner might be untrustworthy, which reduces norm compliance. In other words: the suspicious mindset is conceptualized as the variable that mediates the effect of victim sensitivity on selfish and defensive behavior. However, as described in the introduction of this thesis, research on this mindset and its components is scarce (but see Agroskin et al., 2015; Gerlach et al., 2012; Maltese et al., 2016; Rothmund et al., 2011) and has focused on one component, namely the attributional bias toward malevolence. The other two components – legitimizing cognitions and avoidance motivation – have been neglected so far. In addition, previous studies have mainly used scenario-based approaches with low ecological validity, which is another reason why more direct support for the suspicious mindset is urgently needed. To overcome these limitations, we conducted a study that used a fully immersive virtual environment (c.f. McCall & Blascovich, 2009) to simultaneously investigate all three elements of the suspicious mindset under controlled but ecologically realistic conditions.

In more detail, the study consisted of two parts. Subjects ($n=84$) first filled out an online questionnaire, which included the victim sensitivity scale (Schmitt et al., 2010). Several days or weeks later, they were invited to the virtual reality lab where they solved different tasks with an alleged partner who in reality was a female confederate. To activate a state of suspiciousness, participants were confronted with several untrustworthiness cues in the

beginning¹¹. During the tasks in the virtual world, the (minimum) distance between both players was continuously tracked as a behavioral measure for avoidance motivation (for example, we measured the shortest distance participants kept to the confederate's avatar while reading a number from its back; see also Bailenson et al., 2003). Importantly, all movements of the confederate's avatar were pre-programmed and thus standardized. Afterwards, participants played a trust game (Berg et al., 1995)¹² with their alleged partner over real money and answered some follow-up questions regarding their partner's behavior and intentions. With these follow-up items, we assessed hostile interpretations (e.g., "Do you think the other person has hostile motives?") as well as legitimizing cognitions (e.g., "Do you think it is justified not wanting to work with the other person?") as the other two elements of a suspicious mindset. Trusting behavior in the trust game functioned as our dependent variable (DV).

Contrary to our hypotheses, we were unable to show that approach-avoidance behavior, hostile interpretations, or legitimizing cognitions (as elements of a suspicious mindset) explained, i.e., mediated, the relationship between victim sensitivity and trusting behavior (for an overview of the multiple mediation results, see Figure 2). In fact, VS was totally unrelated to behavioral distrust in our results, as well as to social distancing and to legitimization. However, two findings were noteworthy in this regard: first, although participants high in VS did not withdraw their trust and cooperation in the trust game, they still reported to be more anxious of being exploited by their partner. Secondly, victim sensitivity was associated with an attributional bias, that is, with a heightened readiness to interpret the confederate's behavior in a more hostile way. Thus, victim-sensitive participants were more suspicious of their

¹¹ First, the confederate arrived late to raise doubts about her character and intentions. Second, participants never saw their partner in the real world; and in the virtual world, the confederate was represented by a female avatar with a neutral facial expression (importantly, past research suggests that neutral expressions raise suspicions in victim-sensitive individuals; see Gollwitzer et al., 2012). Third, the possibility of later exploitation in a trust game was made salient before starting the tasks in the virtual world.

¹² In a trust game, a sender/trustor has the opportunity to send points (or money, lottery tickets, etc.) to a receiver/trustee. Points that are sent are usually tripled or quadrupled by the experimenter. Next, the receiver can decide to either keep the tripled/quadrupled points or to share points with the sender. Sending points is thus a risky choice: senders can maximize their profit if receivers share points, but their trust can also be violated.

partner's intentions, which replicates previous findings (Agroskin et al., 2015; Gerlach et al., 2012; Maltese et al., 2016; Rothmund et al., 2011). However, these suspicions did not trigger self-protective behavior. One explanation for this may be that the untrustworthiness cues that we confronted participants with were too subtle to actually justify uncooperative and mistrustful reactions toward the other person. For this reason, future studies should investigate the suspicious mindset and its components in designs containing stronger malevolence cues.

In sum, Study II examined *how exactly* victim sensitivity translates into dysfunctional social behavior, like the withdrawal of trust and cooperation. Although we were unable to find support for most of our hypotheses, Study II still provided some important insights into the social-cognitive processes underlying victim-sensitive people's defensiveness. Most importantly, the results corroborate the notion that VS promotes hostile information processing – and thus, a distorted perception of the social environment – even if only slight evidence for mean intentions is encountered.

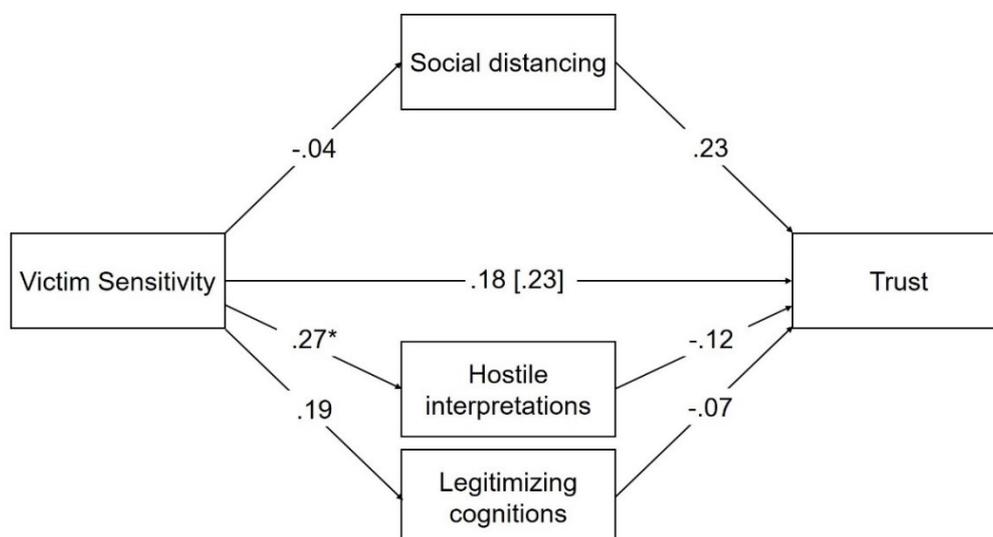


Figure 2. Multiple mediation of the effect of VS on trust via social distancing, hostile interpretations, and legitimizing cognitions, when controlling for general trust (Yamagishi & Yamagishi, 1994), observer sensitivity (Schmitt et al., 2010), neuroticism, and extraversion (Rammstedt & John, 2007). Standardized regression coefficients are depicted.

* $p < .05$ (two-tailed).

Study III

To investigate *why exactly* victim-sensitive individuals tend to behave defensively when the risk for exploitation is considered to be high, Study III was conducted. Two explanations seemed plausible in this context. First, being exploited can be considered a painful experience that has been linked to aversive emotions such as regret, as well as to self-blame and self-recrimination (Vohs et al., 2007). More specifically, Vohs and colleagues argue that being taken advantage of suggests that one has placed one's trust in the wrong person, which carries the threatening implication that one is gullible or stupid. Thus, it could be that individuals high in VS are motivated to react in self-protective and uncooperative ways because the prospect of being exploited threatens their positive self-image. However, being exploited also implies that a disadvantageous outcome is inflicted on the victim by another person, which we argue threatens the victim's *need for control* (e.g., Fiske, 2003; Kay et al., 2008; Rothbaum et al., 1982; Skinner, 1995, 1996). For this reason, it is also possible that victim-sensitive people act pre-emptively defensive because they try to maintain a sense of control over what is happening in their social environment¹³. Notably, if the second alternative were true, then establishing a sense of control – but not affirming a positive self-image – should effectively alleviate victim-sensitive people's tendency to act defensively and selfishly after encountering untrustworthiness cues in a social dilemma situation.

To test these two competing hypotheses against each other, we employed an experimental design in which participants ($n=273$) played a trust game with an alleged (female) partner from another university (for this version of the trust game, see Kuwabara, 2005). Before playing the game, the “partner” allegedly introduced herself with a picture (taken from the Amsterdam Dynamic Facial Expression Set; van der Schalk et al., 2011) and a message (see

¹³ This line of reasoning is strengthened by the fact that victim sensitivity has been associated with a strong motivation to be in control (e.g., Schmitt et al., 1995).

Parks et al., 1996), which were supposed to activate a suspicious mindset (neutral facial expression and competitive message) or to prevent the induction of suspiciousness (joyful expression and cooperative message). Next, participants were either control-affirmed, self-affirmed, or not affirmed at all¹⁴. Finally, participants played the trust game – the sender decision functioned as our DV – and filled out several trait measures, including the victim sensitivity scale (Schmitt et al., 2010).

To test the hypothesized suspiciousness \times affirmation \times victim sensitivity three-way interaction effect, we analyzed the data using a moderated regression model and conducted simple slope analyses in a second step. Consistent with previous research (e.g., Gollwitzer & Rothmund, 2011; Rothmund et al., 2011), results of the moderated regression illustrated that participants high (vs. low) in VS showed less trust in the trust game when they had previously been confronted with untrustworthiness cues (and when no affirmation occurred), but not when the induction of a suspicious mindset had been prevented. Even more importantly, this tendency to distrust others in the presence of untrustworthiness cues was *specifically* alleviated when participants were control-affirmed (see Figure 3). Stated differently: only restoring their sense of control – but not boosting their self-regard – made victim-sensitive participants as trusting as their victim-insensitive counterparts.

In conclusion, a large number of findings has shown that individuals high in VS engage in socially dysfunctional behavior to avoid being taken advantage of by others. However, although victim-sensitive people often act in accordance with their anxious expectation of being exploited, Study III demonstrated that they are able to overcome their habitual distrust if a sense of control is re-established.

¹⁴ In the control-affirmation condition, participants had to recall a recent event in which they experienced a feeling of control (see Kay et al., 2008). In the self-affirmation condition, subjects were asked to describe a recent event in which they had the opportunity to demonstrate a virtue or a talent that was relevant for them (see Monin et al., 2008). In the no-affirmation condition, participants were instructed to list their activities during a typical weekday.

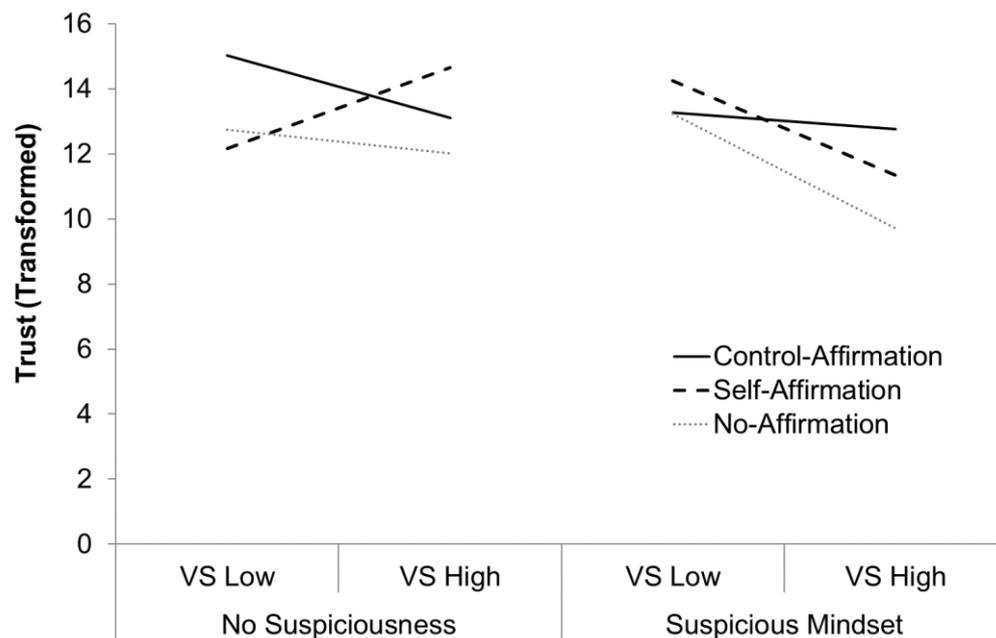


Figure 3. Predicted means visualizing the suspiciousness \times affirmation interaction effect for participants low vs. high in victim sensitivity ($\pm 1 SD$ around the sample mean), when controlling for observer sensitivity (Schmitt et al., 2010) and general trust (Yamagishi & Yamagishi, 1994). Since the trust variable in its original metric was highly skewed, the DV was transformed (see Box & Cox, 1964) before estimating the regression model. Suspiciousness conditions as well as affirmation conditions were contrast-coded.

2.3 Victim Sensitivity and Violations of (Un)Trustworthiness Expectations

Buchholz, M., Bergmann, N., Schubö, A., & Gollwitzer, M. (submitted). Victim sensitivity predicts attention allocation towards violations of untrustworthiness expectancies. Manuscript submitted for publication in *Social Justice Research*.

The results of Studies II and III as well as of previous research suggest that victim-sensitive individuals process information in a way that is consistent with their negative social expectations: people high in VS are asymmetrically sensitive to cues indicating meanness (e.g., Gollwitzer et al., 2009; Gollwitzer et al., 2012), they frequently adopt a suspicious mindset and attribute ulterior motives to other people (e.g., Agroskin et al., 2015; Gerlach et al., 2012; Maltese et al., 2016), and they tend to underestimate their interaction partners' cooperativeness

(Gollwitzer et al., 2012). However, much less is known about how victim-sensitive people react to social information that is *inconsistent* with their initial expectancies about other people's trustworthiness¹⁵. Therefore, the goal of Study IV was to examine how exactly VS shapes the processing of violations of (un)trustworthiness expectations.

Study IV

Since being taken advantage of is highly aversive for individuals high in VS, it could be hypothesized that such individuals are especially receptive to information suggesting that an interaction partner is not as trustworthy as previously expected (i.e., to violations of trustworthiness expectations). However, according to the SeMI model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013), victim sensitivity is not only characterized by an anxious expectation of being exploited, but also by a strong latent motivation to trust other people. Information implying that someone is actually trustworthy, although one expected him or her to be untrustworthy (i.e., information contradicting untrustworthiness expectations), should be more relevant with regard to this need to trust – and might therefore be processed preferentially.

First evidence for this second assumption was provided by Süssenbach and colleagues (2016). In two studies, these authors showed that victim sensitivity predicted enhanced memory accuracy for violations (vs. confirmations) of untrustworthiness expectations, but not for violations of trustworthiness expectations. In addition, victim-sensitive participants also updated their perceptions of other people more strongly when these people showed behavior that contradicted (vs. confirmed) untrustworthiness expectations than when they showed behavior that was inconsistent with trustworthiness expectations. These findings hence supported the notion that violations of untrustworthiness expectations are more motivationally

¹⁵ In this thesis, the terms *expectancy* and *expectation* are used interchangeably. However, the term “expectation” often denotes a more explicit, verbalized belief about the future, whereas an “expectancy” is more frequently understood as something implicit that a person is not fully aware of (e.g., Rief et al., 2015).

relevant for victim-sensitive individuals, and therefore, result in preferential processing. However, Süssenbach et al. did not investigate more basic cognitive processes, such as selective visual attention, which are known to be strongly influenced by a person's motivational state. Because of limited capacities of the perceptual system, selective visual attention prioritizes the processing of those stimuli that are most relevant for a person's goals and needs (e.g., Brosch & van Bavel, 2012; Dietze & Knowles, 2016; Lang et al., 1997; Lavie & Dalton, 2014; Summerfield & Egnér, 2009). Stated differently, if a specific need (e.g., the need to trust) is threatened – and therefore becomes salient – then attention is allocated preferentially to cues that correspond to this need (e.g., cues signaling trustworthiness) and signal a potential opportunity to satisfy that need (c.f. DeWall et al., 2009). Therefore, a more direct test of the assumption that violations of untrustworthiness expectations are particularly motivationally relevant for individuals high in VS would be to show that victim sensitivity predicts attention allocation to these expectancy violations. To test this hypothesis, we employed an eye tracking study, thereby taking advantage of the fact that eye movements are tightly coupled with selective attention processes (Chun et al., 2011; Deubel & Schneider, 1996).

To induce expectations of trustworthiness and untrustworthiness, participants in Study IV ($n=69$) saw computer-generated male faces that were either trustworthy or untrustworthy in appearance (see Oosterhof & Todorov, 2008). In each trial, these faces were accompanied by four words, one of which either violated or confirmed the respective expectation (the remaining words were neutral with regard to trustworthiness). More specifically, faces were always complemented by one (un)trustworthy and three neutral words that were matched with regard to word length and word frequency and that described person characteristics¹⁶. To make sure

¹⁶ The words were taken from two databases (Berlin Affective Word List Reloaded, Võ et al., 2009; Age-Dependent Evaluations of German Adjectives, Grünh & Smith, 2008) and were pretested on trustworthiness. We also implemented a “competition condition”, which entailed both a trustworthy and an untrustworthy word (plus two neutral words) at the same time. However, this competition condition is not referred to further because results were inconsistent; presumably because the conflicting words cancelled each other out.

that participants read the presented words, we implemented a memory task which asked participants if a shown word was old or new after every fifth trial (for an overview of the study procedure, see Figure 4). After the main experiment, victim sensitivity (Schmitt et al., 2010) and other personality traits were assessed in a questionnaire.

To assess visual attention allocation, we recorded participants' eye movements and calculated several eye tracking parameters (e.g., Holmqvist et al., 2011; Inhoff & Radach, 1998; Rayner, 1998; Süßenbach et al., 2012). More specifically, we measured how often participants fixated the respective word of interest, how long participants fixated the word in total (dwell time) and when first landing on it (first fixation duration), and whether the word was fixated first in the trial (destination of the first saccade). Notably, these parameters differently represent either earlier/faster (destination of the first saccade, first fixation duration) or later/slower (dwell time, fixation count) attentional processes (see Holmqvist et al., 2011).

To test our hypothesis that victim sensitivity would predict attention allocation specifically toward violations of untrustworthiness expectations, we used two different approaches. First, we calculated "attentional bias scores" across trials by subtracting means of face-congruent words from means of face-incongruent words. Positive scores therefore reflected preferential attention toward expectancy violations (relative to expectancy confirmations). These difference scores were calculated separately for all four DVs and for both face conditions, and were then correlated with victim sensitivity. Although we focused mainly on these correlation analyses, we additionally used mixed models (i.e., multilevel modeling) to analyze the eye tracking data without aggregating over trials first; however, this was only possible for three of the four DVs. In these multilevel analyses, interaction effects between face type, word type, and victim sensitivity were tested for significance.

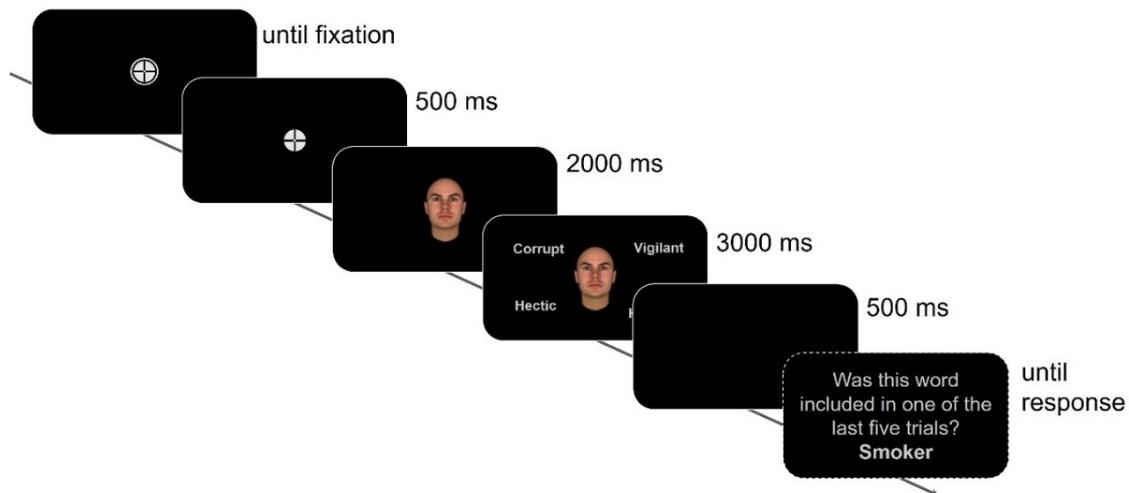


Figure 4. Schematic overview of the experimental paradigm. Each trial started with a fixation dot at screen center that was replaced by a trustworthy- or untrustworthy-looking face when participants successfully fixated the dot. The face was shown for 2000 ms and was then complemented by four words, which were presented for 3000 ms in the four quadrants of the display (so that each word had the same distance to the face). Afterwards, a blank screen appeared, before either the next trial or the memory task started. In total, participants were presented with 300 trials, but faces and words were used repeatedly over trials.

Results showed no effects on the destination of the first saccade measure¹⁷, but we found significant correlations with regard to the other three DVs. As expected, we found that VS was positively correlated with the attentional bias score in first fixation duration in the untrustworthy face condition, but not in the trustworthy face condition. Therefore, victim sensitivity predicted preferential attention in the form of a prolonged first fixation only toward words violating (vs. confirming) untrustworthiness expectations, but not toward words violating trustworthiness expectations. We also discovered significant correlations between VS and difference scores in fixation count and dwell time. However, the correlation with dwell time differences was only significant in the trustworthy face condition and implied an attentional bias associated with *low* victim sensitivity that disappeared with increasing victim sensitivity (see also Figure 5). More specifically, while participants low in victim sensitivity

¹⁷ This was probably due to the fact that participants showed a tendency to fixate the words in the order that corresponded to usual reading habits in Latin script (i.e., from top left to bottom right).

attended longer to trustworthy words confirming the trustworthy facial expression, participants high in VS showed no such consistency or positivity effect. The correlations with fixation count differences were significant in both face conditions and also illustrated that participants low in VS attended preferentially to trustworthy words in general (i.e., in both face conditions), while participants high in VS fixated both word types about equally often. Therefore, the effects in fixation count and dwell time reflected a positivity bias associated with low VS, but no attentional bias toward expectancy violations associated with high VS.

In conclusion, (high) victim sensitivity was associated with an attentional bias toward violations of *untrustworthiness* expectations in our findings, but this bias was only visible in first fixation durations (and therefore in an early stage of attentional processing)¹⁸. In contrast, victim sensitivity did not predict preferential attention toward violations of *trustworthiness* expectations. These results thus corroborate our assumption that information that violates negative initial expectations is more motivationally relevant for victim-sensitive individuals, and therefore receives deeper processing. This may in fact be good news: although their information processing is heavily influenced by their negative social expectations, Study IV showed that individuals high in victim sensitivity do not simply allocate preferential attention to untrustworthiness cues. Instead, trustworthiness-related information that violates these negative expectations is prioritized by selective attention processes – which may reduce victim-sensitive people’s latent fear of exploitation in the long run.

¹⁸ Victim sensitivity did not predict longer dwell times and even fewer fixations on words violating untrustworthiness expectancies. However, the reduced number of fixations may simply be explained by the fact that participants had no time to come back to these words after the prolonged first fixation (because word presentation was limited to 3000 ms). This reasoning is supported by the finding that the number of fixations was negatively related to first fixation durations in our study.

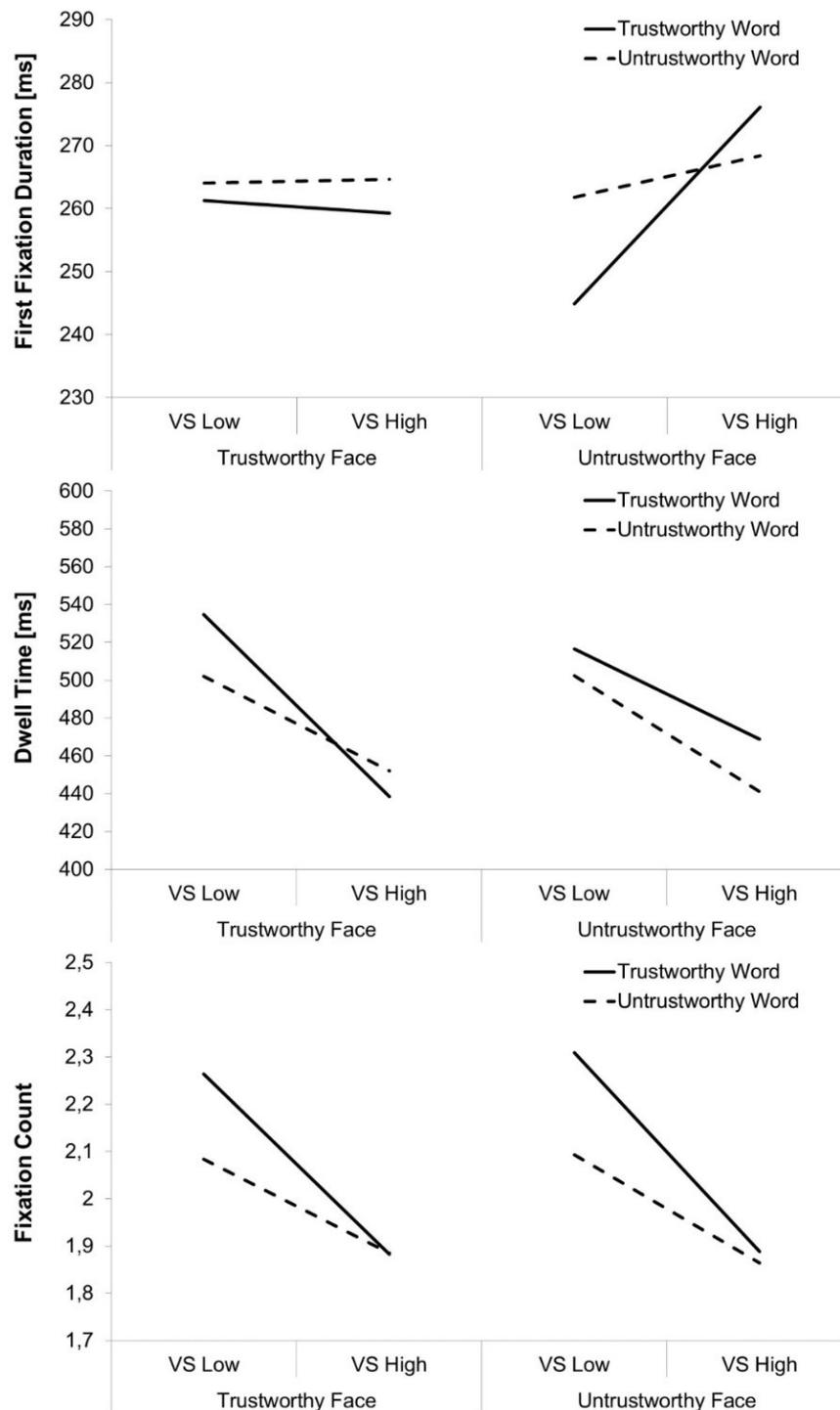


Figure 5. Predicted means in first fixation duration, dwell time, and fixation count (obtained from the multilevel analyses) visualizing the face type \times word type interaction effects for participants low vs. high in victim sensitivity (± 2 *SD* from the sample mean). To account for the nested data structure, random intercepts were modeled for participants.

3 DISCUSSION

In the present thesis, four studies were conducted to further investigate the justice-related personality disposition *victim sensitivity*. To shed light on the consequences of VS as well as on the underlying social-cognitive processes, I relied on new and promising methodological and empirical approaches that have rarely been used in social justice research. For example, in Study I, alignment optimization (Asparouhov & Muthén, 2014) was employed to test for approximate measurement invariance of the justice sensitivity scales and to meaningfully compare mean scores on victim sensitivity between BPD patients and matched healthy controls. In Study II, virtual reality technology (e.g., McCall & Blascovich, 2009) was used to investigate all three elements of the suspicious mindset in a controlled, but ecologically realistic environment. Study III employed a control- and a self-affirmation procedure (Kay et al., 2008; Monin et al., 2008) to examine the motivational processes underlying victim-sensitive people's dysfunctional behavior. Finally, Study IV applied eye tracking (e.g., Deubel & Schneider, 1996; Holmqvist et al., 2011) to measure biases in attentional processing associated with VS.

3.1 Integration of Findings and Directions for Future Research

In the following, I will discuss the contribution of each study to the current state of research on victim sensitivity and the SeMI model, and at the same time, I will address open questions that should be tackled in future studies.

Study I

Study I further investigated the link between victim sensitivity and borderline personality disorder. More precisely, the aim of this study was to follow up on results by Lis et al. (2018), who found higher victim sensitivity in persons with elevated BPD features and

therefore preliminary evidence for the assumption that VS might be an important characteristic of BPD psychopathology. To replicate this finding in a sample of diagnosed BPD patients, Study I examined group differences in victim sensitivity by using the new and promising alignment procedure (Asparouhov & Muthén, 2014).

Two important conclusions could be drawn from this study. First, we established approximate measurement invariance of the victim sensitivity scale between BPD patients and a sample of matched healthy controls, which illustrates that VS can be comparably assessed across persons with and without BPD symptomatology. Second, the results of the alignment procedure demonstrated true-score differences in latent means, with higher VS levels in the BPD patients compared to the control group. As a consequence, individuals with BPD symptoms should show more intense reactions to unjust incidents (such as heightened anger) and they should also be more inclined to retaliate against perpetrators of injustice. Together with the finding that VS partially mediated the relationship between BPD features and self-reported aggression (see Lis et al., 2018), these results lend further support for the assumption that self-related justice concerns may contribute to the interpersonal dysfunction of persons with BPD. However, although our data suggests that victim sensitivity adds to the emergence and maintenance of conduct problems typically associated with BPD, no conclusive inferences on causal relationships can yet be drawn. Longitudinal studies are therefore needed to replicate our cross-sectional findings.

In sum, Study I found additional evidence for the notion that victim sensitivity is a risk factor for, or, at least, a correlate of BPD. Notably, VS has also been shown to be higher in patients with ADHD (Bondü & Esser, 2015) and has been associated with the stabilization of depressive symptoms (Bondü et al., 2017). Put differently: sensitivity to injustice from a victim's perspective might play an important role in the context of other mental disorders as

well. More research investigating victim sensitivity in the context of clinical psychology and psychopathology is therefore clearly warranted.

Study II

In Studies II and III, social-cognitive and motivational processes underlying victim sensitivity were examined in more detail. More specifically, Study II tested the “suspicious mindset hypothesis” derived from the SeMI model to determine *how exactly* victim sensitivity translates into uncooperative and dysfunctional behavior. Similar to previous research on the suspicious mindset (Agroskin et al., 2015; Gerlach et al., 2012; Maltese et al., 2016; Rothmund et al., 2011), we found evidence for a link between VS and an attributional bias toward malevolence. In contrast, we found no support for the other two elements that are assumed to constitute the suspicious mindset, as approach-avoidance behavior and legitimizing cognitions were unrelated to victim sensitivity. Moreover, we were unable to show that any of these variables mediated the effect of VS on trusting behavior. However, this may have been due to the fact that we failed to replicate the established correlation between victim sensitivity and social distrust in our data.

Although these results did not support our hypotheses for the most part, there are a number of possible explanations for this lack of findings. For example, the SeMI model as well as empirical evidence (e.g., Gollwitzer & Rothmund, 2011; Rothmund et al., 2011; Süssenbach & Gollwitzer, 2015) suggest that victim sensitivity predicts uncooperative behavior only if individuals have reason to believe that an interaction partner harbors mean intentions. One reason why we did not find any direct or indirect effects of VS on trusting behavior may therefore be that the untrustworthiness cues that our participants encountered were too subtle to trigger a strong enough state of suspiciousness. Put differently, more evidence for ulterior motives might have been needed to actually justify antisocial behavior. In addition, the SeMI model posits that certain buffers, such as strong norms, can prevent the withdrawal of trust and

cooperation even if a suspicious mindset has been sufficiently activated (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). For example, Dunning et al. (2014) argue that people often trust others because of a moral norm that requires them to show respect for their interaction partner's character – even though they do not actually believe in the other person's benevolence. Importantly, participants in Study II might have felt especially pressured to follow such a norm of respect because (1) they already interacted with their “partner” during the tasks in the virtual world and because (2) they may have expected to face the confederate again after the trust game (e.g. for debriefing).

In sum, Study II found only partial support for the suspicious mindset hypothesis. Although we assessed all three components of this mindset simultaneously, our results merely showed a significant correlation between victim sensitivity and hostile interpretations. As a consequence, there is still a lack of empirical evidence for the two other elements, avoidance motivation and legitimization. Future research should thus investigate these two components more thoroughly, and should also explore whether the suspicious mindset may actually be better described by additional or different elements. In addition, more theoretical and empirical work should be devoted to boundary conditions for the effect of suspiciousness on uncooperative and distrustful behavior. Potential buffering variables (such as strong norms) are only briefly discussed in the SeMI model, and – to the best of my knowledge – have not yet been systematically defined and studied. More research on these buffers is therefore urgently needed.

Study III

Following up on Study II, Study III examined *why exactly* victim-sensitive people tend to act uncooperatively in social situations, with the goal to find effective ways to alleviate such negative behavior. Based on theoretical considerations, two different motivations seemed plausible in this context: first, victim-sensitive people might act selfishly, or rather defensively,

because anticipated exploitation threatens their positive self-image (Vohs et al., 2007). Second, it could be that people high in VS act in socially dysfunctional ways because they try to maintain a sense of control over their social environment. We expected the second assumption to be true, which is why we predicted that restoring a sense of control – but not boosting their self-regard – would effectively alleviate victim-sensitive people’s antisocial behavior.

To compare the two competing hypotheses against each other, Study III used an experimental between-subjects design in which half of the participants were confronted with untrustworthiness cues, before they were either self-affirmed, control-affirmed, or not affirmed at all. In line with our hypothesis, our results demonstrated that (1) victim sensitivity predicted social mistrust after a suspicious mindset had been activated and that (2) this tendency to distrust others was mitigated specifically when participants were control-affirmed. These results thus corroborate the notion that when expecting exploitation, individuals high in VS react in self-protective ways to avoid a loss of (perceived) control.

Notably, participants in Study III were confronted with strong cues of untrustworthiness (i.e., a neutral facial expression along with a competitive message) to activate a strong state of suspiciousness with the goal of testing our hypotheses as rigorously and as conservatively as possible. The fact that restoring a sense of control still made victim-sensitive participants as trusting as their victim-insensitive counterparts therefore proves that this is an effective way to alleviate adverse behavioral consequences of VS. However, Study III only tested immediate effects of the control-affirmation procedure on participants’ behavior in a trust game. It is therefore unclear how long-lasting these effects are. In fact, it seems questionable whether a low-threshold intervention such as recalling a recent event in which one experienced a sense of control would produce long-term effects. Thus, although Study III provides a good starting point for explaining what motivates dysfunctional behavior in victim-sensitive persons, more work needs to be done in this regard. Most importantly, the results of the present research

should be used to develop interventions capable of permanently modifying dysfunctional behavior patterns.

Study IV

In Study IV, we investigated the attentional processing of violations of (un)trustworthiness expectancies as a function of victim sensitivity. More specifically, we tested the hypothesis that violations of untrustworthiness expectations (e.g., information implying that an untrustworthy-looking person is actually trustworthy) are more *motivationally relevant* for victim-sensitive individuals than violations of trustworthiness expectations (e.g., information implying that a person is less trustworthy than expected), and therefore receive preferential attention. As such, Study IV also examined the previously untested assumption of the SeMI model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013) that VS reflects not only an anxious expectation of being exploited, but also a latent motivation to trust others.

To disentangle the different aspects of victim sensitivity, Study IV used an experimental within-subjects design in which participants' eye movements were recorded. In line with our hypothesis and previous findings (Süssenbach et al., 2016), our data showed that VS predicted an attentional bias toward violations of untrustworthiness expectations, but not toward violations of trustworthiness expectations. Therefore, the results of Study IV corroborate the assumption that information that contradicts negative initial expectations is particularly motivationally relevant for victim-sensitive people. This, in turn, provides a first indication that victim sensitivity is indeed characterized by a latent motivation to trust others.

However, these findings represent only rather indirect support for this hypothesis. Therefore, stronger and more direct evidence is needed to be able to conclude with certainty that victim sensitivity can be conceptualized as a combination of a latent expectation regarding others malevolence together with a strong need to trust others. Future studies should therefore use empirical and methodological approaches with which the underlying psychological

processes can be mapped more directly. For example, physiological measures such as EEG might be helpful in this regard. In addition, future research might want to manipulate participants' motivational state in a more salient way. In our study, we used facial expressions to induce expectations of untrustworthiness and therefore, to threaten victim-sensitive people's need to trust. However, future studies might benefit from using designs in which participants are not mere observers, but actually experience victimization to be able to draw more conclusive inferences about their motivational state.

3.2 Methodological Limitations

The four studies described in this thesis further corroborate the SeMI model's predictions and expand our knowledge about the social-cognitive processes underlying victim sensitivity, but there are also some methodological limitations to consider. First and foremost, showing a significant relationship once in a single study can never provide conclusive evidence for or against a particular hypothesis (Ioannidis, 2005). Therefore, a replication of the results presented in this thesis is urgently needed, especially in light of the recent replicability crisis in (social) psychology and other social sciences (e.g., Camerer et al., 2016; Camerer et al., 2018; Ebersole et al., 2016; Klein et al., 2018; Open Science Collaboration, 2015). This replicability crisis has revealed that a significant portion of research discoveries cannot be replicated, in particular those showing interaction effects (such as investigated in Studies III and IV). One conclusion that has been drawn from this crisis is that transparency in research needs to be increased, for example by pre-registering hypotheses, methods and analyses plans before collecting any data (van 't Veer & Giner-Sorolla, 2016). Pre-registration can help to create more robust research findings because it clearly distinguishes confirmatory (i.e., a priori specified) from exploratory (i.e., post hoc determined) analyses and reduces reporting and publication bias – that is, the selective reporting and publishing of only significant results.

Since none of the studies described in this thesis were pre-registered, it is particularly important that pre-registration takes place in future replications to strengthen the credibility of our findings.

In addition, sample sizes were rather small in some of the studies, which further calls for a replication of findings. For example, Study II used a sample of only 84 participants to investigate the mediating effects of all three elements that are assumed to constitute the suspicious mindset. In our results, none of the hypothesized indirect effects reached significance. However, Monte Carlo simulations illustrated that in multiple mediation models, much larger sample sizes (i.e., around 400 to 500 participants) are needed to detect small to medium sized indirect effects with a power of 80% (Fritz & MacKinnon, 2007; Ma & Zeng, 2014; Thoemmes et al., 2010). Therefore, future studies should investigate the suspicious mindset and its components in better powered studies. Similarly, Study III with its sample size of 273 was underpowered for studying the expected interaction between victim sensitivity, suspiciousness, and affirmation condition at an alpha level of 5%. According to an a priori power analysis, we would have needed at least 315 participants to detect a small- to medium-sized three-way interaction effect with a power of 80%, if $\alpha=5\%$ (calculated with G*Power 3.1; Faul et al., 2009). To still ensure sufficient power in our sample, we had to raise the alpha level to 10% – thereby increasing the possibility for detecting false-positive results.

Another constraint lies in the fact that Studies II to IV were conducted mainly with undergraduate students, which may raise concerns about the generalizability of findings. It can be assumed that students are younger and better educated than the population average. However, research using representative data has shown that demographic effects in the justice sensitivity facets are small, with factors such as gender, age, and education explaining only little variance in victim sensitivity (Schmitt et al., 2010). It is therefore unlikely that the effects described in this thesis are influenced significantly by age or education. Nevertheless, future

research replicating the present findings should aim to use more heterogeneous samples to see whether similar patterns of results can be observed.

3.3 Conclusion

The present thesis makes important contributions to research on victim sensitivity by expanding our knowledge about its role in clinical contexts (Study I) and the underlying social-cognitive and motivational processes (Studies II through IV). Overall, two important conclusions can be drawn from this research. First, the four studies illustrate once again that victim sensitivity shapes behavior and information processing in a dysfunctional way. Study I, for example, presented further evidence for the assumption that VS contributes to emotional and behavioral problems of person with BPD. In addition, Study II replicated the finding that VS is associated with an attributional bias toward malevolence – and therefore, with a distorted perception of the social environment.

However, the present thesis also demonstrates that it is possible to alleviate these adverse consequences of victim sensitivity, and therefore, to improve social interactions. Studies III and IV show promising results in this regard. For example, Study III illustrated that victim-sensitive people are able to overcome their habitual suspiciousness if a sense of control is reinforced. Moreover, Study IV demonstrated that violations of untrustworthiness expectations are more deeply processed by victim-sensitive people, which in the long run may reduce these negative expectations, and therefore, their latent fear of exploitation.

Nevertheless, the findings presented in this thesis have to be seen as tentative evidence that needs to be replicated in high-powered and, ideally, pre-registered designs with representative samples.

REFERENCES

- Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology, 67*(5), 422-436. <https://doi.org/10.1037/h0040968>
- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299). Academic Press.
- Agroskin, D., Jonas, E., & Traut-Mattausch, E. (2015). When suspicious minds go political: Distrusting and justifying the system at the same time. *Political Psychology, 36*(6), 613-629. <https://doi.org/10.1111/pops.12185>
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.).
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508. <https://doi.org/10.1080/10705511.2014.919210>
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin, 29*(7), 819-833. <https://doi.org/10.1177/0146167203029007002>
- Ball, J. S., & Links, P. S. (2009). Borderline personality disorder and childhood trauma: Evidence for a causal relationship. *Current Psychiatry Reports, 11*, 63-68. <https://doi.org/10.1007/s11920-009-0010-4>
- Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45-103). Erlbaum.
- Bateman, A. W., & Krawitz, R. (2013). *Borderline personality disorder: An evidence-based guide for generalist mental health professionals*. Oxford University Press.

- Battle, C. L., Shea, M. T., Johnson, D. M., Yen, S., Zlotnick, C., Zanarini, M. C., Sanislow, C. A., Skodol, A. E., Gunderson, J. G., Grilo, C. M., McGlashan, T. H., & Morey, L. C. (2004). Childhood maltreatment associated with adult personality disorders: Findings from the Collaborative Longitudinal Personality Disorders Study. *Journal of Personality Disorders, 18*(2), 193-211. <https://doi.org/10.1521/pedi.18.2.193.32777>
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*(2), 243-267. <https://doi.org/10.1037/0033-2909.115.2.243>
- Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., & Rammstedt, B. (2014). Measuring four perspectives of justice sensitivity with two items each. *Journal of Personality Assessment, 96*(3), 380-390. <https://doi.org/10.1080/00223891.2013.836526>
- Baumert, A., Nazlic, T., & Alrich, K. (2014). Konflikte bearbeiten durch Relativierung von Gerechtigkeitsvorstellungen [Resolving conflicts by qualifying subjective justice views]. *Konfliktdynamik, 1*, 36-43. <https://doi.org/10.5771/2193-0147-2014-1-36>
- Baumert, A., Otto, K., Thomas, N., Bobocel, D. R., & Schmitt, M. (2012). Processing of unjust and just information: Interpretation and memory performance related to dispositional victim sensitivity. *European Journal of Personality, 26*(2), 99-110. <https://doi.org/10.1002/per.1844>
- Baumert, A., Rothmund, T., Thomas, N., Gollwitzer, M., & Schmitt, M. (2013). Justice as a moral motive: Belief in a just world and justice sensitivity as potential indicators of the justice motive. In K. Heinrichs, F. K. Oser, & T. Lovat (Eds.), *Handbook of moral motivation: Theories, models, applications* (pp. 159-180). Sense Publishers.

- Baumert, A., & Schmitt, M. (2016). Justice sensitivity. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 161-180). Springer.
https://doi.org/10.1007/978-1-4939-3216-0_9
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122-142. <https://doi.org/10.1006/game.1995.1027>
- Bondü, R., & Elsner, B. (2015). Justice sensitivity in childhood and adolescence. *Social Development*, *24*(2), 420-441. <https://doi.org/10.1111/sode.12098>
- Bondü, R., & Esser, G. (2015). Justice and rejection sensitivity in children and adolescents with ADHD symptoms. *European Child & Adolescent Psychiatry*, *24*, 185-198.
<https://doi.org/10.1007/s00787-014-0560-9>
- Bondü, R., & Krahé, B. (2015). Links of justice and rejection sensitivity with aggression in childhood and adolescence. *Aggressive Behavior*, *41*(4), 353-368.
<https://doi.org/10.1002/ab.21556>
- Bondü, R., & Richter, P. (2016). Linking forms and functions of aggression in adults to justice and rejection sensitivity. *Psychology of Violence*, *6*(2), 292-302.
<https://doi.org/10.1037/a0039200>
- Bondü, R., Sahyazici-Knaak, F., & Esser, G. (2017). Long-term associations of justice sensitivity, rejection sensitivity, and depressive symptoms in children and adolescents. *Frontiers in Psychology*, *8*, 1446. <https://doi.org/10.3389/fpsyg.2017.01446>
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B*, *26*(2), 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Brosch, T., & Van Bavel, J. J. (2012). The flexibility of emotional attention: Accessible social identities guide rapid attentional orienting. *Cognition*, *125*(2), 309-316.
<https://doi.org/10.1016/j.cognition.2012.07.007>

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Byrne, B. M., & van de Vijver, F. J. R. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema, 29*(4), 539-551. <https://doi.org/10.7334/psicothema2017.178>
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmedj, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433-1436. <https://doi.org/10.1126/science.aaf0918>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*, 637-644. <https://doi.org/10.1038/s41562-018-0399-z>
- Carroll, A. (2009). Are you looking at me? Understanding and managing paranoid personality disorder. *Advances in Psychiatric Treatment, 15*(1), 40-48. <https://doi.org/10.1192/apt.bp.107.005421>
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology, 62*, 73-101. <https://doi.org/10.1146/annurev.psych.093008.100427>
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). Oxford University Press.

- Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 584-627). John Wiley & Sons, Inc.
- Crosby, F. (1976). A model of egoistical relative deprivation. *Psychological Review*, *83*(2), 85-113. <https://doi.org/10.1037/0033-295X.83.2.85>
- Dalbert, C. (2001). *The justice motive as a personal resource: Dealing with challenges and critical life events*. Kluwer Academic/Plenum Publishers.
- Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, *35*(10), 867-881. <https://doi.org/10.1037/0003-066X.35.10.867>
- Davidov, E. (2010). Testing for comparability of human values across countries and time with the third round of the European Social Survey. *International Journal of Comparative Sociology*, *51*(3), 171-191. <https://doi.org/10.1177/0020715210363534>
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169-193. <https://doi.org/10.1146/annurev.ps.31.020180.001125>
- De Cremer, D., Snyder, M., & Dewitte, S. (2001). 'The less I trust, the less I contribute (or not)?' The effects of trust, accountability and self-monitoring in social dilemmas. *European Journal of Social Psychology*, *31*(1), 93-107. <https://doi.org/10.1002/ejsp.34>
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827-1837. [https://doi.org/10.1016/0042-6989\(95\)00294-4](https://doi.org/10.1016/0042-6989(95)00294-4)
- DeWall, C. N., Maner, J. K., & Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: Selective attention to signs of acceptance. *Journal of Personality and Social Psychology*, *96*(4), 729-741. <https://doi.org/10.1037/a0014634>

- Dietze, P., & Knowles, E. D. (2016). Social class and the motivational relevance of other human beings: Evidence from visual attention. *Psychological Science, 27*(11), 1-11. <https://doi.org/10.1177/0956797616667721>
- Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122-141. <http://doi.org/10.1037/a0036673>
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., ... Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67*, 68-82. <https://doi.org/10.1016/j.jesp.2015.10.012>
- Elovainio, M., Kivimäki, M., & Vahtera, J. (2002). Organizational justice: Evidence of a new psychosocial predictor of health. *American Journal of Public Health, 92*(19), 105-108. <https://doi.org/10.2105/AJPH.92.1.105>
- Emmons, R. A. (1989). Exploring the relations between motives and traits: The case of narcissism. In D. M. Buss & N. Cantor (Eds.), *Personality Psychology: Recent trends and emerging directions* (pp. 32-44). Springer. https://doi.org/10.1007/978-1-4684-0634-4_3
- Faccenda, L., Pantaléon, N., & Reynes, E. (2009). Significant predictors of soccer players' moral functioning from components of contextual injustice, sensitivity to injustice and moral atmosphere. *Social Justice Research, 22*, 399-415. <https://doi.org/10.1007/s11211-009-0105-4>

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences, 36*(5), 1015-1029. [https://doi.org/10.1016/S0191-8869\(03\)00197-1](https://doi.org/10.1016/S0191-8869(03)00197-1)
- Fiske, S. T. (2003). Five core social motives, plus or minus five. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated social perception: The Ontario Symposium* (Vol. 9, pp. 233-246). Erlbaum.
- Fiske, S. T. (2018). *Social beings: Core motives in social psychology* (4th ed.). John Wiley & Sons, Inc.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233-239. <https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Gerlach, T. M., Allemand, M., Agroskin, D., & Denissen, J. J. A. (2012). Justice sensitivity and forgiveness in close interpersonal relationships: The mediating role of mistrustful, legitimizing, and pro-relationship cognitions. *Journal of Personality, 80*(5), 1373-1413. <https://doi.org/10.1111/j.1467-6494.2012.00762.x>
- Gollwitzer, M., & Rothmund, T. (2009). When the need to trust results in unethical behavior: The Sensitivity to Mean Intentions (SeMI) model. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 135-152). Information Age Publishing.
- Gollwitzer, M., & Rothmund, T. (2011). What exactly are victim-sensitive persons sensitive to? *Journal of Research in Personality, 45*(5), 448-455. <https://doi.org/10.1016/j.jrp.2011.05.003>

- Gollwitzer, M., Rothmund, T., Alt, B., & Jekel, M. (2012). Victim sensitivity and the accuracy of social judgments. *Personality and Social Psychology Bulletin, 38*(8), 975-984. <https://doi.org/10.1177/0146167212440887>
- Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality, 43*(6), 999-1005. <https://doi.org/10.1016/j.jrp.2009.07.003>
- Gollwitzer, M., Rothmund, T., & Süßenbach, P. (2013). The Sensitivity to Mean Intentions (SeMI) model: Basic assumptions, recent findings, and potential avenues for future research. *Social and Personality Psychology Compass, 7*(7), 415-426. <https://doi.org/10.1111/spc3.12041>
- Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical effects of justice sensitivity perspectives on prosocial and antisocial behavior. *Social Justice Research, 18*(2), 183-201. <https://doi.org/10.1007/s11211-005-7368-1>
- Gollwitzer, M., Süßenbach, P., & Hannuschke, M. (2015). Victimization experiences and the stabilization of victim sensitivity. *Frontiers in Psychology, 6*, 439. <https://doi.org/10.3389/fpsyg.2015.00439>
- Gollwitzer, M., & van Prooijen, J.-W. (2016). Psychology of justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 61-82). Springer. https://doi.org/10.1007/978-1-4939-3216-0_4
- Greenberg, J. (2006). Losing sleep over organizational injustice: Attenuating insomniac reactions to underpayment inequity with supervisory training in interactional justice. *Journal of Applied Psychology, 91*(1), 58-69. <https://doi.org/10.1037/0021-9010.91.1.58>
- Greenberg, J. (2010). Organizational injustice as an occupational health risk. *Academy of Management Annals, 4*(1), 205-243. <https://doi.org/10.5465/19416520.2010.481174>

- Grühn, D., & Smith, J. (2008). Characteristics for 200 words rated by young and older adults: Age-dependent evaluations of German adjectives (AGE). *Behavior Research Methods, 40*(4), 1088-1097. <https://doi.org/10.3758/BRM.40.4.1088>
- Herman, J. L., Perry, J. C., & van der Kolk, B. A. (1989). Childhood trauma in borderline personality disorder. *The American Journal of Psychiatry, 146*(4), 490-495. <https://doi.org/10.1176/ajp.146.4.490>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Huseman, R. C., Hatfield, J. D., & Miles, E. W. (1987). A new perspective on equity theory: The equity sensitivity construct. *Academy of Management Review, 12*(2), 222-234. <https://doi.org/10.5465/amr.1987.4307799>
- Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-53). Elsevier.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin, 37*(5), 701-713. <https://doi.org/10.1177/0146167211400208>
- Kay, A. C., Gaucher, D., Napier, J. L., Callan, M. J., & Laurin, K. (2008). God and the government: Testing a compensatory control mechanism for the support of external systems. *Journal of Personality and Social Psychology, 95*(1), 18-35. <https://doi.org/10.1037/0022-3514.95.1.18>

- Klass, E. T. (1978). Psychological effects of immoral actions: The experimental evidence. *Psychological Bulletin*, 85(4), 756-771. <https://doi.org/10.1037/0033-2909.85.4.756>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443-490. <https://doi.org/10.1177/2515245918810225>
- Kuwabara, K. (2005). Nothing to fear but fear itself: Fear of fear, fear of greed and gender effects in two-person asymmetric social dilemmas. *Social Forces*, 84(2), 1257-1272. <https://doi.org/10.1353/sof.2006.0020>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons & M. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97-135). Erlbaum.
- Lavelle, J. J., Harris, C. M., Rupp, D. E., Herda, D. N., Young, R. F., Hargrove, M. B., Thornton-Lugo, M. A., & McMahan, G. C. (2018). Multifoci effects of injustice on counterproductive work behaviors and the moderating roles of symbolization and victim sensitivity. *Journal of Organizational Behavior*, 39(8), 1022-1039. <https://doi.org/10.1002/job.2280>
- Lavie, N., & Dalton, P. (2014). Load theory of attention and cognitive control. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 56-75). Oxford University Press.

- Lazarus, S. A., Cheavens, J. S., Festa, F., & Rosenthal, M. Z. (2014). Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clinical Psychology Review, 34*(3), 193-205.
<https://doi.org/10.1016/j.cpr.2014.01.007>
- Lerner, M. J. (1977). The justice motive: Some hypotheses as to its origins and forms. *Journal of Personality, 45*(1), 1-52. <https://doi.org/10.1111/j.1467-6494.1977.tb00591.x>
- Lerner, M. J. (1980). *The belief in a just world. A fundamental delusion*. Plenum Press.
- Lis, S., Schaedler, A., Liebke, L., Hauschild, S., Thome, J., Schmahl, C., Stahlberg, D., Kleindienst, N., & Bohus, M. (2018). Borderline personality disorder features and sensitivity to injustice. *Journal of Personality Disorders, 32*(2), 192-206.
https://doi.org/10.1521/pedi_2017_31_292
- Lotz, S., Okimoto, T. G., Schlösser, T., & Fetchenhauer, D. (2011). Punitive versus compensatory reactions to injustice: Emotional antecedents to third-party interventions. *Journal of Experimental Social Psychology, 47*(2), 477-480.
<https://doi.org/10.1016/j.jesp.2010.10.004>
- Lotz, S., Schlösser, T., Cain, D. M., & Fetchenhauer, D. (2013). The (in)stability of social preferences: Using justice sensitivity to predict when altruism collapses. *Journal of Economic Behavior & Organization, 93*, 141-148.
<https://doi.org/10.1016/j.jebo.2013.07.012>
- Lovas, L., & Wolt, R. (2002). Sensitivity to injustice in the context of some personality traits. *Studia Psychologica, 44*, 125-131.
- Ma, Z.-w., & Zeng, W.-n. (2014). A multiple mediator model: Power analysis based on Monte Carlo simulation. *American Journal of Applied Psychology, 3*(3), 72-79.
<https://doi.org/10.11648/j.ajap.20140303.15>

- Maltese, S., Baumert, A., Schmitt, M. J., & MacLeod, C. (2016). How victim sensitivity leads to uncooperative behavior via expectancies of injustice. *Frontiers in Psychology, 6*, 2059. <https://doi.org/10.3389/fpsyg.2015.02059>
- Mancke, F., Herpertz, S. C., & Bertsch, K. (2015). Aggression in borderline personality disorder: A multidimensional model. *Personality Disorders: Theory, Research, and Treatment, 6*(3), 278-291. <https://doi.org/10.1037/per0000098>
- McCall, C., & Blascovich, J. (2009). How, when, and why to use digital experimental virtual environments to study social behavior. *Social and Personality Psychology Compass, 3*(5), 744-758. <https://doi.org/10.1111/j.1751-9004.2009.00195.x>
- Mikula, G. (2002). Perspective-related differences in interpretations of injustice in close relationships. In C. F. Graumann & W. Kallmeyer (Eds.), *Perspective and perspectivation in discourse* (pp. 251-262). John Benjamins Publishing Company.
- Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. *Personality and Social Psychology Bulletin, 24*(7), 769-783. <https://doi.org/10.1177/0146167298247009>
- Mohiyeddini, C., & Schmitt, M. J. (1997). Sensitivity to befallen injustice and reactions to unfair treatment in a laboratory situation. *Social Justice Research, 10*(3), 333-353. <https://doi.org/10.1007/BF02683307>
- Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*(1), 76-93. <https://doi.org/10.1037/0022-3514.95.1.76>
- Montada, L. (1993). Understanding oughts by assessing moral reasoning or moral emotions. In G. G. Noam & T. E. Wren (Eds.), *The moral self* (pp. 292-309). MIT Press.

- Montada, L. (2007). Justice conflicts and the justice of conflict resolution. In K. Törnblom & R. Vermunt (Eds.), *Distributive and procedural justice. Research and social applications* (pp. 255-268). Ashgate.
- Montada, L., Schmitt, M., & Dalbert, C. (1986). Thinking about justice and dealing with one's own privileges: A study of existential guilt. In H. W. Bierhoff, R. L. Cohen, & J. Greenberg (Eds.), *Justice in social relations* (pp. 125-143). Plenum Press.
- Montada, L., & Schneider, A. (1989). Justice and emotional reactions to the disadvantaged. *Social Justice Research*, 3(4), 313-344. <https://doi.org/10.1007/BF01048081>
- Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, 5, 978. <https://doi.org/10.3389/fpsyg.2014.00978>
- Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby effect. *Emotion*, 9(1), 118-122. <https://doi.org/10.1037/a0014540>
- Ogata, S. N., Silk, K. R., Goodrich, S., Lohr, N. E., Westen, D., & Hill, E. M. (1990). Childhood sexual and physical abuse in adult patients with borderline personality disorder. *The American Journal of Psychiatry*, 147(8), 1008-1013.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA*, 105(32), 11087-11092. <https://doi.org/10.1073/pnas.0805664105>
- Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, 9(1), 128-133. <https://doi.org/10.1037/a0014520>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

- Parks, C. D., Henager, R. F., & Scamahorn, S. D. (1996). Trust and reactions to messages of intent in social dilemmas. *Journal of Conflict Resolution*, *40*(1), 134-151.
<https://doi.org/10.1177/0022002796040001007>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203-212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubö, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry*, *28*(5), 378-38. <https://doi.org/10.1097/YCO.0000000000000184>
- Rimé, B., Philippot, P., Boca, S., & Mesquita, B. (1992). Long-lasting cognitive and social consequences of emotion: Social sharing and rumination. *European Review of Social Psychology*, *3*(1), 225-258. <https://doi.org/10.1080/14792779243000078>
- Rothbaum, F., Weisz, J. R., & Snyder, S. S. (1982). Changing the world and changing the self: A two-process model of perceived control. *Journal of Personality and Social Psychology*, *42*(1), 5-37. <https://doi.org/10.1037/0022-3514.42.1.5>
- Rothmund, T., Gollwitzer, M., & Klimmt, C. (2011). Of virtual victims and victimized virtues: Differential effects of experienced aggression in video games on social cooperation. *Personality and Social Psychology Bulletin*, *37*(1), 107-119.
<https://doi.org/10.1177/0146167210391103>
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651-665. <https://doi.org/10.1111/j.1467-6494.1967.tb01454.x>

- Rubin, Z., & Peplau, A. (1973). Belief in a just world and reactions to another's lot: A study of participants in the national draft lottery. *Journal of Social Issues, 29*(4), 73-93. <https://doi.org/10.1111/j.1540-4560.1973.tb00104.x>
- Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues, 31*(3), 65-89. <https://doi.org/10.1111/j.1540-4560.1975.tb00997.x>
- Schmahl, C., Herpertz, S. C., Bertsch, K., Ende, G., Flor, H., Kirsch, P., Lis, S., Meyer-Lindenberg, A., Rietschel, M., Schneider, M., Spanagel, R., Treede, R.-D., & Bohus, M. (2014). Mechanisms of disturbed emotion processing and social interaction in borderline personality disorder: State of knowledge and research agenda of the German Clinical Research Unit. *Borderline Personality Disorder and Emotion Dysregulation, 1*, 12. <https://doi.org/10.1186/2051-6673-1-12>
- Schmitt, M. (1996). Individual differences in Sensitivity to Befallen Injustice (SBI). *Personality and Individual Differences, 21*(1), 3-20. [https://doi.org/10.1016/0191-8869\(96\)00028-1](https://doi.org/10.1016/0191-8869(96)00028-1)
- Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The justice sensitivity inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*, 211-238. <https://doi.org/10.1007/s11211-010-0115-2>
- Schmitt, M., & Dörfel, M. (1999). Procedural injustice at work, justice sensitivity, job satisfaction and psychosomatic well-being. *European Journal of Social Psychology, 29*(4), 443-453. [https://doi.org/10.1002/\(SICI\)1099-0992\(199906\)29:4<443::AID-EJSP935>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1099-0992(199906)29:4<443::AID-EJSP935>3.0.CO;2-C)
- Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment, 21*(3), 202-211. <https://doi.org/10.1027/1015-5759.21.3.202>

- Schmitt, M. J., & Mohiyeddini, C. (1996). Sensitivity to befallen injustice and reactions to a real-life disadvantage. *Social Justice Research, 9*(3), 223-238.
<https://doi.org/10.1007/BF02197249>
- Schmitt, M. J., Neumann, R., & Montada, L. (1995). Dispositional sensitivity to befallen injustice. *Social Justice Research, 8*(4), 385-407. <https://doi.org/10.1007/BF02334713>
- Skinner, E. A. (1995). *Perceived control, motivation, and coping*. Sage Publications.
- Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology, 71*(3), 549-570. <https://doi.org/10.1037/0022-3514.71.3.549>
- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.
<https://doi.org/10.1086/209528>
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Süssenbach, P., Bohner, G., & Eyssel, F. (2012). Schematic influences of rape myth acceptance on visual information processing: An eye-tracking approach. *Journal of Experimental Social Psychology, 48*(3), 660-668.
<https://doi.org/10.1016/j.jesp.2011.12.005>
- Süssenbach, P., & Gollwitzer, M. (2015). Us(ed): The role of victim sensitivity in potentially exploitative intergroup relationships. *Group Processes & Intergroup Relations, 18*(2), 241-255. <https://doi.org/10.1177/1368430214556700>
- Süssenbach, P., Gollwitzer, M., Mieth, L., Buchner, A., & Bell, R. (2016). Trustworthy tricksters: Violating a negative social expectation affects source memory and person perception when fear of exploitation is high. *Frontiers in Psychology, 7*, 1-12.
<https://doi.org/10.3389/fpsyg.2016.02037>

- Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling, 17*(3), 510-534. <https://doi.org/10.1080/10705511.2010.489379>
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. In A. Kingstone & M. B. Miller (Eds.), *The year in cognitive neuroscience 2008: Annals of the New York Academy of Sciences* (Vol. 1124., p. 208-224). Blackwell Publishing.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813-833. <https://doi.org/10.1521/soco.2009.27.6.813>
- van den Bos, K., Maas, M., Waldring, I. E., & Semin, G. R. (2003). Toward understanding the psychology of reactions to perceived fairness: The role of affect intensity. *Social Justice Research, 16*(2), 151-168. <https://doi.org/10.1023/A:1024252104717>
- van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expressions Set (ADFES). *Emotion, 11*(4), 907-920. <https://doi.org/10.1037/a0023853>
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492. <https://doi.org/10.1080/17405629.2012.686740>
- van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology, 6*, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology, 67*, 2-12. <https://doi.org/10.1016/j.jesp.2016.03.004>

- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534-538. <https://doi.org/10.3758/BRM.41.2.534>
- Vohs, K. D., Baumeister, R. F., & Chin, J. (2007). Feeling duped: Emotional, motivational, and cognitive aspects of being exploited by others. *Review of General Psychology*, *11*(2), 127-141. <https://doi.org/10.1037/1089-2680.11.2.127>
- Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin*, *27*(7), 899-909. <https://doi.org/10.1177/0146167201277012>
- Weiss, H. M., Suckow, K., & Cropanzano, R. (1999). Effects of justice conditions on discrete emotions. *Journal of Applied Psychology*, *84*(5), 786-794. <https://doi.org/10.1037/0021-9010.84.5.786>
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129-166. <https://doi.org/10.1007/BF02249397>
- Zanarini, M. C., Williams, A. A., Lewis, R. E., Reich, R. B., Vera, S. C., Marino, M. F., Levin, A., Yong, L., & Frankenburg, F. R. (1997). Reported pathological childhood experiences associated with the development of borderline personality disorder. *The American Journal of Psychiatry*, *154*(8), 1101-1106. <https://doi.org/10.1176/ajp.154.8.1101>
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497-1517. <https://doi.org/10.1111/j.1751-9004.2008.00109.x>

APPENDIX

Manuscript A

Buchholz, M., Neukel, C., Steinmann, S., Lis, S., & Gollwitzer, M. (submitted). Measuring justice sensitivity in patients with borderline personality disorder: Do we really measure the same? Manuscript submitted for publication in *Assessment*.

**Measuring Justice Sensitivity in Patients with Borderline Personality Disorder:
Do We Really Measure The Same?**

Merle Buchholz¹, Corinne Neukel², Sylvia Steinmann^{3,4}, Stefanie Lis^{3,5}, & Mario Gollwitzer⁶

¹Department of Psychology, Philipps-Universität Marburg

²Department of General Psychiatry, Center for Psychosocial Medicine, Heidelberg University

³Medical Faculty Mannheim, Heidelberg University

⁴Department of Psychosomatic Medicine and Psychotherapy, Central Institute of Mental
Health, Mannheim

⁵Institute of Psychiatric and Psychosomatic Psychotherapy, Central Institute of Mental Health,
Mannheim

⁶Department of Psychology, Ludwig-Maximilians-Universität München

Author Note

Stefanie Lis and Mario Gollwitzer contributed equally to this work. This research was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, project number 290878970-GRK 2271, Project 4) to Philipps-Universität Marburg. We have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Merle Buchholz, Department of Psychology, Philipps-Universität Marburg, Gutenbergstraße 18, 35032 Marburg, Germany. Phone: +4964212823792; email: merle.buchholz@uni-marburg.de

Abstract

Recent findings suggest that persons with Borderline Personality Disorder (BPD) features are more sensitive to injustice as assessed by the Justice Sensitivity Inventory (JSI). There are two equally plausible explanations for these findings: one is that observed mean differences reflect true-score differences; the other assumes that persons with BPD use the response scales of the JSI differently than healthy controls (measurement inequality). To compare these two explanations against each other, measurement invariance between 91 BPD patients and 455 matched healthy controls was tested via the alignment procedure. Results demonstrate approximate measurement invariance of the JSI. Latent mean comparisons yielded higher levels of victim, observer, and beneficiary sensitivity in the BPD patient group compared to the healthy controls group; these differences reflect true-score differences. Our findings have implications for the measurement of justice sensitivity in BPD patients and confirm increased justice sensitivity as a feature of BPD in a clinical sample.

[150 words]

Keywords: justice sensitivity inventory, borderline personality disorder, measurement invariance, approximate approach, alignment optimization

Measuring Justice Sensitivity in Patients with Borderline Personality Disorder: Do We Really Measure The Same?

Justice and morality play an important role in the everyday life of most individuals. People do not only want to be treated fairly and equally by others, they also want to see justice prevail in the world. On the one hand, this concern for justice seems to be universal (Montada, 2007), but, on the other hand, research suggests that people differ reliably in how they perceive, evaluate, and react to unjust situations (e.g., Schmitt et al., 2005). Schmitt, Neumann, and Montada (1995) were the first to conceptualize “Justice Sensitivity” (JS) as a personality trait. It captures individual differences regarding emotional and behavioral reactions to experienced, observed, or committed injustice. These individual differences can be measured via self-report scales. The latest version of these scales that measure Justice Sensitivity from the perspectives of a victim, an observer, a beneficiary, and a perpetrator, is now referred to as the Justice Sensitivity Inventory (JSI; Schmitt et al., 2010).

In the last two decades, the JS construct has received considerable attention in social psychology and, more recently, also in clinical psychology and psychopathology. Although a dispositional concern for justice may contribute to psychological well-being and mental health (e.g., Dalbert, 2001), there is also evidence that justice sensitivity may be a risk factor for reduced interpersonal functioning and life satisfaction (Baumert & Schmitt, 2016). For instance, links between justice sensitivity and aggressive behavior were found for both children and adolescents, as well as for adults (Bondü & Krahe, 2015; Bondü & Richter, 2016). Here, especially justice sensitivity from the perspectives of a victim and an observer predicted higher levels of physical, relational, verbal, proactive, and reactive aggression. In addition, (victim) justice sensitivity was shown to be higher in participants with ADHD, and this JS facet also partially mediated the link between ADHD symptoms and comorbid problems like depression, low self-esteem, and maladaptive behavior. Therefore, it is

reasonable to assume that (victim) justice sensitivity promotes conduct problems and internalizing problems in persons with ADHD, presumably through increased dysfunctional thoughts and beliefs and distressing emotions (Bondü & Esser, 2015). Finally, (victim) justice sensitivity was linked to the stabilization of depressive symptoms in children and adolescents in longitudinal data—a finding that further corroborates the notion that justice sensitivity may result in or relate to a broad range of mental health problems (Bondü & Elsner, 2015; Bondü et al., 2017).

Recent research has explored the relation between JS and Borderline Personality Disorder features (e.g., Lis et al., 2018). Borderline Personality Disorder (BPD) is a severe mental disorder with a prevalence of 1.6% in the general population and 20% of the psychiatric, inpatient population (Ellison et al., 2018). It is characterized by affective instability, an unstable self-concept, and interpersonal problems associated to non-suicidal self-injurious behavior, high rates of suicides and suicide attempts, and an increased mortality rate (Schneider et al., 2019). Many diagnosed patients report a history of abuse and victimization during childhood and adolescence (e.g., Ball & Links, 2009; Battle et al., 2004; Zanarini et al., 1997) that may influence the patients' sensitivity for injustice. Prior experiences of injustice in the context of neglect and abuse might lead to a strong sense of justice and to attempts to preserve justice at all costs, which in turn might contribute to the interpersonal dysfunction of people with BPD (Bateman & Krawitz, 2013). Building on this reasoning, Lis and colleagues (2018) found that participants with elevated BPD features had higher mean scores on two subscales of the Justice Sensitivity Inventory—victim sensitivity and observer sensitivity—than subjects presenting moderate or low BPD features. In addition, and similar to the findings of Bondü and Esser (2015), victim sensitivity partially mediated the relationship between BPD features and aggression. These results suggest that JS may be a factor relevant for BPD psychopathology, or, at least, a correlate of BPD.

That said, in clinical research there is always an alternative explanation for mean score differences between healthy participants and participants with mental disorders or an accentuation of specific personality features: possibly, these differences do not reflect “true-score” differences, but rather differences in how patients and healthy controls use the response scales of self-report questionnaires. Applied to our present case, differences between BPD patients and healthy controls on the JSI subscales might not represent “true-score” differences in justice sensitivity. For instance, BPD patients might be more acquiescent than healthy participants, that is, they may tend to prefer more extreme response categories irrespective of the item content (a “response set” that has been initially investigated among anxiety-prone individuals, cf. Lewis & Taylor, 1955). Alternatively, BPD patients may tend to aggravate symptoms that are associated with emotional relevance and moral concerns (for respective effects in clinical samples, see Gollwitzer et al., 2005; Grillo et al., 1994; Nichols & Greene, 1997; Schoenberg et al., 2004). Finally, BPD patients may exhibit a tendency to respond in a “socially desirable” fashion, that is, respond more affirmatively to statements that display a good moral character to others (e.g., Gollwitzer et al., 2005; Paulhus et al., 1995; Paulhus & Reid, 1991). Because the JSI items are emotionally relevant, indicative of a “moral concern”, and at least partly socially desirable, it is quite possible that response styles, but not true-score differences, explain the observed mean differences between participants with BPD symptomatology and healthy controls on the JSI subscales. Notably, this would severely threaten the interpretation that justice sensitivity is a meaningful psychological variable in the context of BPD.

Technically, differences in how people from different subgroups treat a particular response scale can be captured by several methodological approaches, including statistical controlling for covariates (e.g., by controlling for stable “socially desirable responding” tendencies; cf. Paulhus, 1991). This approach requires that (1) the source of a particular

response set can be defined and measured, and that (2) the covariate captures this source (e.g., “social desirability” concerns) reliably and validly. A second approach is to interpret responders’ scores differently depending on their subgroup or their assignment to a “latent class” (e.g., Gollwitzer et al., 2005). This requires that the subgroups/classes can be meaningfully interpreted in terms of response set differences. Each of these approaches has its own strengths and weaknesses, and it is beyond the scope of this paper to go into detail here. More basically, before applying one of these remedial approaches, it makes sense to ask whether there is reason to believe that the measurement model reliably differs between different subgroups in the first place. This is what measurement invariance tests do.

As stated by van de Schoot et al. (2015, p. 1), measurement invariance “requires that the association between the items (or test scores) and the latent factors (or latent traits) of individuals should not depend on group membership or measurement occasion (i.e., time)”. Methodologists have defined three levels of measurement invariance that are typically tested: configural, metric, and scalar invariance (e.g., Brown, 2006; Steenkamp & Baumgartner, 1998). Configural invariance means that the measurement model has the same form or “configuration”; thus, this form of invariance requires that the number of factors and the pattern of factor loadings is equal across groups or measurement occasions. Metric invariance refers to the fact that the measurement metric needs to be the same in all groups or time points, which requires factor loadings to be equivalent. Lastly, scalar invariance demands equality in indicator intercepts and therefore in indicator difficulties as well.

Measurement invariance is a central prerequisite for comparing scores obtained in different groups—such as BPD patients and healthy controls—directly to each other. Group comparisons of latent means, for example, are meaningful only if factor loadings and indicator intercepts are invariant across groups (Brown, 2006; Davidov, 2010; Steenkamp & Baumgartner, 1998). Therefore, using the Justice Sensitivity Inventory to investigate the role

of justice sensitivity in the context of BPD requires showing that the inventory is actually “measurement-invariant”. Most commonly, multi-group confirmatory factor analysis is used to test for measurement equivalence, whereby increasingly restrictive invariance constraints are imposed on the model and, then, evaluated (Brown, 2006; Jöreskog, 1971; Sörbom, 1974). However, this traditional approach has recently been criticized to be overly strict because it requires measurement parameters to be *exactly the same* across groups or time, which is an unrealistic assumption in most cases (van de Schoot et al., 2015). In other words, it is often impossible to achieve complete (scalar) invariance. For this reason, novel approaches rooted in the concept of “approximate measurement invariance” have been developed (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2013; van de Schoot et al., 2013). In contrast to the exact approach, these methods allow researchers to meaningfully compare latent means even in the presence of a certain amount of non-invariance. Stated differently, some slight differences are permitted as long as the measurements of the construct are sufficiently (i.e., approximately) similar.

One method that is based on this concept of approximate equivalence is the alignment optimization technique introduced by Asparouhov and Muthén (2014). This approach estimates trustworthy factor means and variances without constraining factor loadings and indicator intercepts to equality. It does so by incorporating a simplicity function, which minimizes the number of non-invariant parameters and the total amount of non-invariance in a way similar to that of rotation functions in exploratory factor analysis. More specifically, the alignment procedure has two steps: first, a configural model is fitted to the data (Horn & McArdle, 1992). Second, this configural model undergoes an optimization process such that the model fit is retained but the extent of non-invariance minimized. Thus, alignment greatly simplifies and automates the assessment of measurement invariance. However, the results of the alignment optimization should only be trusted and relied upon if less than 25% of all

measurement parameters have been found to be non-invariant (Muthén & Aspharouhov, 2014). In the present study, alignment optimization is applied to test for (approximate) measurement invariance of the Justice Sensitivity Inventory and latent mean differences between BPD patients and healthy controls.

Method

Participants

For the present analysis, we combined data from 91 BPD patients obtained from several studies at the Institute for Psychiatric and Psychosomatic Psychotherapy and the Department for Psychosomatic Medicine and Psychotherapy, CIMH Mannheim. Recruitment for these studies was performed using the research database of the Clinical Research Unit “Mechanisms of disturbed emotion processing in BPD” funded by the German Research Foundation (DFG; KFO 256, Schmahl et al., 2014). All studies were conducted in accordance with the Declaration of Helsinki, and were approved by the Research Ethics Board of the University of Heidelberg. Subjects provided written informed consent prior to study participation. Trained clinical psychologists used the International Personality Disorder Examination (Loranger, 1999) to diagnose BPD in accordance with DSM-IV. Axis I disorders were assessed using the Structured Clinical Interview for DSM-IV (SCID-I; First et al., 1997). All patients met at least five of the nine DSM-IV criteria for BPD. We measured severity of BPD symptoms with the Borderline-Symptom List (BSL-23 (short version); Bohus et al., 2009). The BSL-23 is a self-report measure that assesses severity of borderline-specific symptomatology during the previous week as the average score of 23 items that are rated on a five-point Likert scale. Possible scores range from 0-4 with higher scores indicating a higher degree of BPD psychopathology.

All patients were female. Ages ranged between 18 and 50 years, with a mean age of 31 years ($SD = 8.6$). Of the 91 patients, 2% had eight years of education, 4% nine years, 26%

ten years, and 67% at least twelve years. Mean BSL score was 1.77 ($SD = 0.96$) indicating a high symptom severity (Kleindienst et al., 2020).

Each case in the clinical sample was matched with five non-clinical cases, respectively; matching criteria were age, sex, and years of education. These control cases were taken from four different samples: two representative samples from the German population ($n=107$ and $n=308$, respectively) collected by Schmitt et al. (2010; sample 1) and Maes and Schmitt (1999), and two demographically heterogeneous convenience samples ($n=7$ and $n=33$) collected by Schmitt et al. (2010; sample 2) and Schmitt et al. (2005; Study 2, sample 2). Thus, the matched control sample consisted of $n = 455$ female participants with a mean age of 31 years (range = 18 to 50 years, $SD = 8.5$). Years of education ranged between 8 and 12 years, and the distribution was comparable to the clinical sample: most participants had more than 11 years of education (67% in both samples), and 33% had 10 years of education or less in both samples. Healthy controls did not fill out the BSL-23.

Materials

In all samples, justice sensitivity was assessed with the Justice Sensitivity Inventory (JSI; Schmitt et al., 2010), which measures victim (VS), observer (OS), beneficiary (BS), and perpetrator sensitivity (PS) with ten items each. Items were always rated on six-point Likert scales (0=absolutely disagree to 5=absolutely agree). An English version of the items can be found in Appendix A.

Statistical Analyses

Descriptive statistics for the two groups are reported in Table 1. To test for measurement invariance of the Justice Sensitivity Inventory between BPD patients and healthy controls, we used the alignment procedure described in Asparouhov and Muthén (2014). Mplus version 8.4 with the Combination Add-On was used for all analyses (Muthén & Muthén, 1998-2017). To reduce the number of free model parameters and therefore model

complexity, we generated five item parcels for each JSI subscale by computing the mean of two items. This was necessary because of the moderate sample sizes of $n_1 = 91$ and $n_2 = 455$. As described in Schmitt et al. (2010), item pairs were equal for all scales; that is, parcels always consisted of items 1+6, 2+7, 3+8, 4+9, 5+10. In addition, we excluded perpetrator sensitivity from all analyses because (1) there were many missing values in this subscale among the healthy controls and because (2) this facet shows large overlaps in content and high correlations with beneficiary sensitivity (Schmitt et al., 2010).

The Hypothesized Model

The assumed factor model is shown schematically in Figure 1. This model assumes that for both BPD patients and healthy controls, the JSI (excluding the perpetrator sensitivity scale) is represented by a 3-factor structure comprising of victim, observer, and beneficiary sensitivity. In addition, the model assumes that (1) these three factors are correlated, (2) indicator residuals are uncorrelated, and (3) cross-loadings are zero (i.e., parcels do not have loadings on factors they were not intended to measure).

Results

In accordance with the two-step alignment procedure described above, we first fitted a configural model to the data. However, the hypothesized model (see Figure 1) with uncorrelated error terms and with secondary loadings fixed to zero showed an unsatisfactory model fit, $\chi^2(174) = 690.842, p < .001$; RMSEA = .104; CFI = .929. Inspection of the model modification indices suggested that this misfit was largely due to restricting error terms of parcels consisting of identical item pairs to be uncorrelated across scales¹. For this reason, we tested a model that allowed for such correlated residuals in a second step. With this adjustment, model fit increased significantly, $\chi^2(144) = 429.114, p < .001$; RMSEA = .085; CFI = .961). According to standard fit criteria (e.g., van de Schoot et al., 2012), the Root Mean Square Error of Approximation (RMSEA) of the modified configural model suggests a

mediocre fit, while the CFI suggests an adequate to good fit. Importantly, the final aligned model has the same fit as this baseline model.

Non-Invariance Results

Based on the modified configural model with correlated error terms, the alignment optimization was applied. Because we tested for measurement invariance across two groups, fixed alignment based on Maximum Likelihood estimation was used (Asparouhov & Muthén, 2014). In fixed alignment, one group (here: the healthy controls group) serves as the reference group in which factor means are fixed to zero.

Measurement invariance results for each measurement parameter are summarized in Table 2. According to the fixed alignment optimization output, all factor loadings as well as all intercepts were approximately invariant between the groups (the average invariance index was 0.775). However, some of the parameters still showed a higher degree of non-invariance. For example, when reviewing the results in Table 2 it becomes apparent that the intercept of Parcel V4 exhibited a large contribution to the simplicity function, resulting in the highest overall contribution (-1.051). Because larger contributions to the fitting function reflect higher levels of non-invariance (Asparouhov & Muthén, 2014), this result suggests that this parcel was the least invariant among all indicator variables. In addition, four factor loadings (V4, V5, B4, and B5) showed a very low R^2 value, which can be interpreted as an invariance index as well. More specifically, the R^2 measure indicates for each parameter how much variation across groups can be explained by variation in factor means and variances; as such, values close to 1 suggest a high degree of invariance and values close to 0 a low degree of invariance (Asparouhov & Muthén, 2014). Thus, five measurement parameters (i.e., 17% in total) showed evidence for some degree of non-invariance, which might have reached significance if sample sizes had been larger. However, even if considering these parameters as somewhat less invariant, the amount of non-invariant parameters is far below the cut-off value of 25%

proposed as a rule of thumb by Muthén and Asparouhov (2014). Therefore, latent mean estimates derived from the alignment results can be considered trustworthy and reliable.

Factor Mean Results

Table 3 presents the latent factor means as estimated by the fixed alignment method for both groups. Mean values for the healthy controls are zero because this group functioned as the reference group. As can be seen, the factor means of all three justice sensitivity facets were significantly higher in the BPD patients group compared to the healthy controls group.

Discussion

Justice Sensitivity (JS) has been proposed as an important feature of BPD as suggested by recent research comparing JS scores between persons with different levels of BPD features (Lis et al., 2018). However, so far, this assumption has not been confirmed in a clinical sample with a BPD diagnosis based on the evaluation of diagnostic criteria in standardized clinical interviews by trained clinicians. Moreover, unless measurement equivalence between patients and control participants is established, observed mean differences may either reflect true-score differences or response-set differences. Using the relatively new and promising alignment method (Asparouhov & Muthén, 2014), we were able to demonstrate approximate measurement invariance of the Justice Sensitivity Inventory (JSI) across BPD patients and matched healthy controls. More specifically, we found none of the factor loadings or indicator intercepts to be significantly non-invariant, which ensures comparability of the estimated factor means for victim, observer, and beneficiary sensitivity. This comparison showed significant group differences in all facet means, supporting the interpretation of higher concerns about justice in people with BPD (Lis et al., 2018).

Interestingly, in our analysis we found beneficiary sensitivity to be higher in the BPD patients group than in the healthy controls group—a difference that did not reach significance in the previous study by Lis et al. (2018). There might be at least three reasons for this

discrepancy: first, Lis et al. compared participants high, moderate, and low in BPD features (i.e., subclinical samples), while our analyses compared healthy participants with a clinical sample of BPD patients diagnosed according to diagnostic criteria of the DSM by trained psychologists. Thus, it may be that only a higher level of BPD symptomatology results in higher beneficiary sensitivity scores. Second, the discrepant results might simply be rooted in the different methodological approaches used, as latent group comparisons (used in the present study) have more statistical power than repeated-measures ANOVAs (used in Lis et al., 2018) due to corrections for measurement error (c.f. Brown, 2006). Third, sample sizes were larger in our study, which should have increased the power to detect mean differences even more. Nevertheless, future research should address this discrepant finding to see whether the difference in beneficiary sensitivity found in our study can be replicated.

In sum, our study corroborates the notion that justice sensitivity is a meaningful psychological construct in the context of BPD, and that by using the JSI, victim, observer, and beneficiary sensitivity can be assessed comparably across individuals with and without BPD psychopathology. This has important implications for clinical research investigating the relevance of heightened justice concerns for affective instability, and interpersonal dysfunction in BPD, and will therefore contribute to our understanding of this severe personality disorder in the long run.

Endnotes

1. The subscales of the Justice Sensitivity Inventory are designed to describe the same types of injustice and to have similar wording across perspectives, which may be considered a common source of variance. Thus, error correlations among parcels consisting of analogue items might imply that people judge specific kinds of unjust events as more (or less) morally wrongful, regardless of the perspective from which the injustice is experienced (Schmitt et al., 2010).

References

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495-508.
<https://doi.org/10.1080/10705511.2014.919210>
- Ball, J. S., & Links, P. S. (2009). Borderline personality disorder and childhood trauma: Evidence for a causal relationship. *Current Psychiatry Reports*, *11*, 63-68.
<https://doi.org/10.1007/s11920-009-0010-4>
- Bateman, A. W., & Krawitz, R. (2013). *Borderline personality disorder: An evidence-based guide for generalist mental health professionals*. Oxford University Press.
- Battle, C. L., Shea, M. T., Johnson, D. M., Yen, S., Zlotnick, C., Zanarini, M. C., Sanislow, C. A., Skodol, A. E., Gunderson, J. G., Grilo, C. M., McGlashan, T. H., & Morey, L. C. (2004). Childhood maltreatment associated with adult personality disorders: Findings from the Collaborative Longitudinal Personality Disorders Study. *Journal of Personality Disorders*, *18*(2), 193-211. <https://doi.org/10.1521/pedi.18.2.193.32777>
- Baumert, A., & Schmitt, M. (2016). Justice sensitivity. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 161-180). Springer.
https://doi.org/10.1007/978-1-4939-3216-0_9
- Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R.-D., Domsalla, M., Chapman, A. L., Steil, R., Philipsen, A., & Wolf, M. (2009). The short version of the Borderline Symptom List (BSL-23): Development and initial data on psychometric properties. *Psychopathology*, *42*(1), 32-39. <https://doi.org/10.1159/000173701>
- Bondü, R., & Elsner, B. (2015). Justice sensitivity in childhood and adolescence. *Social Development*, *24*(2), 420-441. <https://doi.org/10.1111/sode.12098>
- Bondü, R., & Esser, G. (2015). Justice and rejection sensitivity in children and adolescents

- with ADHD symptoms. *European Child & Adolescent Psychiatry*, 24, 185-198.
<https://doi.org/10.1007/s00787-014-0560-9>
- Bondü, R., & Krahe, B. (2015). Links of justice and rejection sensitivity with aggression in childhood and adolescence. *Aggressive Behavior*, 41(4), 353-368.
<https://doi.org/10.1002/ab.21556>
- Bondü, R., & Richter, P. (2016). Linking forms and functions of aggression in adults to justice and rejection sensitivity. *Psychology of Violence*, 6(2), 292-302.
<https://doi.org/10.1037/a0039200>
- Bondü, R., Sahyazici-Knaak, F., & Esser, G. (2017). Long-term associations of justice sensitivity, rejection sensitivity, and depressive symptoms in children and adolescents. *Frontiers in Psychology*, 8, 1446. <https://doi.org/10.3389/fpsyg.2017.01446>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Dalbert, C. (2001). *The justice motive as a personal resource: Dealing with challenges and critical life events*. Springer. <https://doi.org/10.1007/978-1-4757-3383-9>
- Davidov, E. (2010). Testing for comparability of human values across countries and time with the third round of the European Social Survey. *International Journal of Comparative Sociology*, 51(3), 171-191. <https://doi.org/10.1177/0020715210363534>
- Ellison, W. D., Rosenstein, L. K., Morgan, T. A., & Zimmerman, M. (2018). Community and clinical epidemiology of borderline personality disorder. *Psychiatric Clinics of North America*, 41(4), 561-573. <https://doi.org/10.1016/j.psc.2018.07.008>
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., & Benjamin, L. S. (1997). *User's guide for the structured clinical interview for DSM-IV Axis I disorders (SCID-I) - clinical version*. American Psychiatric Press.
- Gollwitzer, M., Eid, M., & Jürgensen, R. (2005). Response styles in the assessment of anger

- expression. *Psychological Assessment*, 17(1), 56-69. <https://doi.org/10.1037/1040-3590.17.1.56>
- Grillo, J., Brown, R. S., Hilsabeck, R., Price, J. R., & Lees-Haley, P. R. (1994). Raising doubts about claims of malingering: Implications of relationships between MCMI-II and MMPI-2 performances. *Journal of Clinical Psychology*, 50(4), 651-655. [https://doi.org/10.1002/1097-4679\(199407\)50:4<651::AID-JCLP2270500424>3.0.CO;2-C](https://doi.org/10.1002/1097-4679(199407)50:4<651::AID-JCLP2270500424>3.0.CO;2-C)
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117-144, <https://doi.org/10.1080/03610739208253916>
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426.
- Kleindienst, N., Jungkunz, M., & Bohus, M. (2020). A proposed severity classification of borderline symptoms using the borderline symptom list (BSL-23). *Borderline Personality Disorder and Emotion Dysregulation*, 7, 11. <https://doi.org/10.1186/s40479-020-00126-6>
- Lewis, N. A., & Taylor, J. A. (1955). Anxiety and extreme response preferences. *Educational and Psychological Measurement*, 15(2), 111-116. <https://doi.org/10.1177/001316445501500203>
- Lis, S., Schaedler, A., Liebke, L., Hauschild, S., Thome, J., Schmahl, C., Stahlberg, D., Kleindienst, N., & Bohus, M. (2018). Borderline personality disorder features and sensitivity to injustice. *Journal of Personality Disorders*, 32(2), 192-206. https://doi.org/10.1521/pedi_2017_31_292
- Loranger, A. W. (1999). *International Personality Disorder Examination (IPDE): DSM-IV and ICD-10 modules*. Psychological Assessment Resources.

- Maes, J., & Schmitt, M. (1999). More on ultimate and immanent justice: Results from the research project "Justice as a Problem within Reunified Germany". *Social Justice Research, 12*(2), 65-78. <https://doi.org/10.1023/A:1022039624976>
- Montada, L. (2007). Justice conflicts and the justice of conflict resolution. In K. Törnblom & R. Vermunt (Eds.), *Distributive and procedural justice. Research and social applications* (pp. 255-268). Ashgate.
- Muthén, B., & Asparouhov, T. (2013). *BSEM Measurement Invariance Analysis. Mplus Web Notes: No. 17*. Available online at: www.statmodel.com
- Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology, 5*, 978. <https://doi.org/10.3389/fpsyg.2014.00978>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment, 68*(2), 251-266. https://doi.org/10.1207/s15327752jpa6802_3
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press. <https://doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*(2), 100-108. <https://doi.org/10.1177/0146167295212001>
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*(2), 307-317. <https://doi.org/10.1037/0022-3514.60.2.307>

- Schmahl, C., Herpertz, S. C., Bertsch, K., Ende, G., Flor, H., Kirsch, P., Lis, S., Meyer-Lindenberg, A., Rietschel, M., Schneider, M., Spanagel, R., Treede, R.-D., & Bohus, M. (2014). Mechanisms of disturbed emotion processing and social interaction in borderline personality disorder: State of knowledge and research agenda of the German Clinical Research Unit. *Borderline Personality Disorder and Emotion Dysregulation, 1*, 3-17. <http://dx.doi.org/10.1186/2051-6673-1-12>
- Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*, 211-238. <https://doi.org/10.1007/s11211-010-0115-2>
- Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment, 21*, 202-211. <https://doi.org/10.1027/1015-5759.21.3.202>
- Schmitt, M. J., Neumann, R., & Montada, L. (1995). Dispositional sensitivity to befallen injustice. *Social Justice Research, 8*, 385-407. <https://doi.org/10.1007/BF02334713>
- Schneider, F., Erhart, M., Hewer, W., Loeffler, L. A. K., & Jacobi, F. (2019). Mortality and medical comorbidity in the severely mentally ill—a German registry study. *Deutsches Ärzteblatt, 116*(23-24), 405-411.
- Schoenberg, M. R., Dorr, D., Morgan, C. D., & Burke, M. (2004). A comparison of the MCMI-III personality disorder and modifier indices with the MMPI-2 clinical and validity scales. *Journal of Personality Assessment, 82*(3), 273-280. https://doi.org/10.1207/s15327752jpa8203_03
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*(2), 229-239. <https://doi.org/10.1111/j.2044-8317.1974.tb00543.x>

- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90.
<https://doi.org/10.1086/209528>
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 770.
<https://doi.org/10.3389/fpsyg.2013.00770>
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.
<https://doi.org/10.1080/17405629.2012.686740>
- van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, *6*, 1064.
<https://doi.org/10.3389/fpsyg.2015.01064>
- Zanarini, M. C., Williams, A. A., Lewis, R. E., Reich, R. B., Vera, S. C., Marino, M. F., Levin, A., Yong, L., & Frankenburg, F. R. (1997). Reported pathological childhood experiences associated with the development of borderline personality disorder. *American Journal of Psychiatry*, *154*(8), 1101– 1106.
<https://doi.org/10.1176/ajp.154.8.1101>

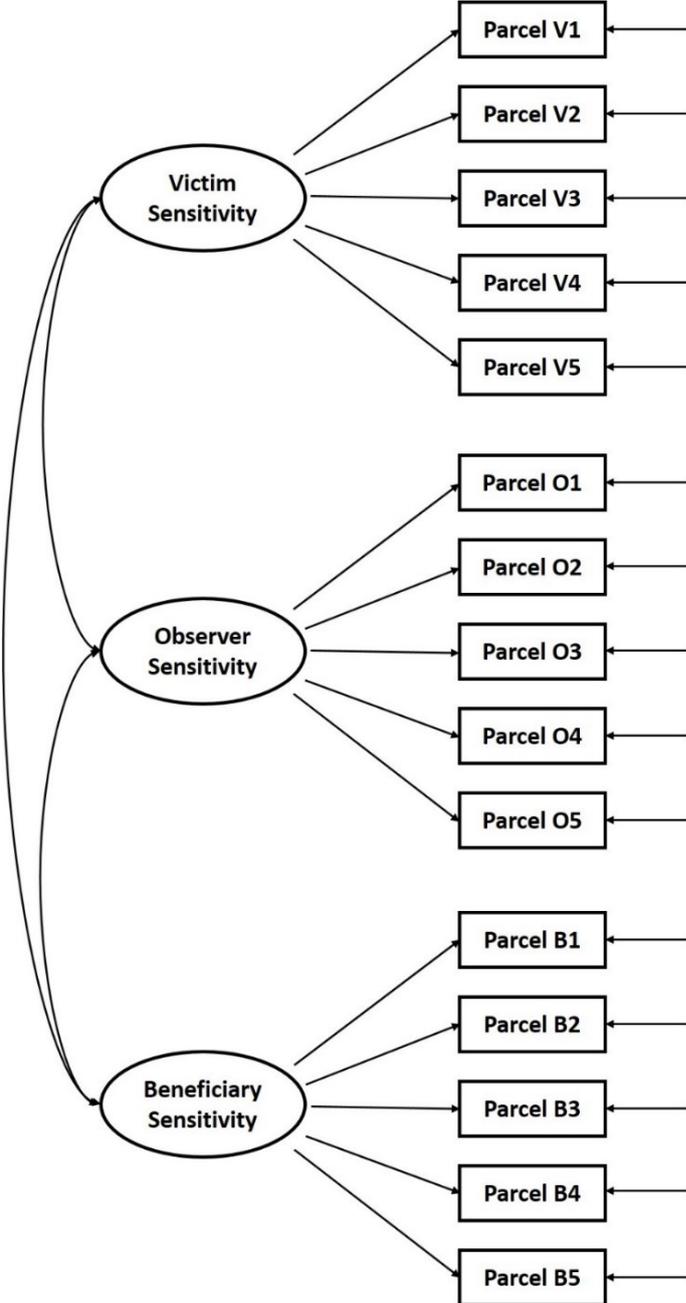


Figure 1. Three-factor model of justice sensitivity.

Table 1*Subsample Characteristics*

Subgroup	Age	VS	OS	BS	PS
Healthy controls (<i>n</i> =455)	31 (8.5)	2.75 (1.05)	2.78 (0.99)	2.73 (1.11)	3.04 (1.48)
BPD patients (<i>n</i> =91)	31 (8.6)	3.36 (1.08)	3.27 (1.12)	3.21 (1.17)	3.79 (1.14)

Note. Means and standard deviations (in parentheses). VS = victim sensitivity; OS = observer sensitivity; BS = beneficiary sensitivity; PS = perpetrator sensitivity.

Table 2*Invariance Results and Fit Statistics*

Variable	Factor loading			Indicator intercept			Parcel Total contri- bution
	Approx. invariant across groups	R^2	Fit function contri- bution	Approx. invariant across groups	R^2	Fit function contri- bution	
VS							
Parcel V1	yes	0.534	-0.380	yes	0.766	-0.488	-0.867
Parcel V2	yes	0.707	-0.339	yes	0.994	-0.339	-0.678
Parcel V3	yes	0.967	-0.317	yes	0.969	-0.394	-0.712
Parcel V4	yes	0.000	-0.387	yes	0.833	-0.663	-1.051
Parcel V5	yes	0.000	-0.532	yes	0.980	-0.316	-0.849
OS							
Parcel O1	yes	0.935	-0.328	yes	0.985	-0.343	-0.671
Parcel O2	yes	0.981	-0.319	yes	0.996	-0.324	-0.643
Parcel O3	yes	0.757	-0.327	yes	0.996	-0.320	-0.647
Parcel O4	yes	0.987	-0.318	yes	0.965	-0.383	-0.701
Parcel O5	yes	0.787	-0.326	yes	0.995	-0.331	-0.657
BS							
Parcel B1	yes	0.593	-0.366	yes	0.996	-0.317	-0.683
Parcel B2	yes	0.564	-0.375	yes	0.971	-0.362	-0.737
Parcel B3	yes	1.000	-0.316	yes	1.000	-0.317	-0.633
Parcel B4	yes	0.000	-0.471	yes	0.988	-0.374	-0.844
Parcel B5	yes	0.000	-0.365	yes	0.999	-0.316	-0.682

Note. Approximate measurement (non-)invariance of intercepts and factor loadings. Fit function contributions represent the contribution made by each parameter (or parcel) to the final simplicity function (the smaller the contribution, the more invariant). R^2 values represent the degree of invariance (0 = low degree of invariance, 1 = high degree of invariance). VS = victim sensitivity, OS = observer sensitivity, BS = beneficiary sensitivity.

Table 3*Factor Mean Comparison*

Factor	Healthy controls	BPD patients
Victim sensitivity	0.000	0.586*
Observer sensitivity	0.000	0.514*
Beneficiary sensitivity	0.000	0.449*

* Significant difference in factor means across groups ($p < .05$)

Appendix A**Table A1***Victim Sensitivity Items*

Nr.	Item
1	It bothers me when others receive something that ought to be mine
2	It makes me angry when others receive a reward that I have earned
3	I cannot easily bear it when others profit unilaterally from me
4	It takes me a long time to forget when I have to fix others' carelessness
5	It gets me down when I get fewer opportunities than others to develop my skills
6	It makes me angry when others are undeservingly better off than me
7	It worries me when I have to work hard for things that come easily to others
8	I ruminate for a long time when other people are treated better than me
9	It burdens me to be criticized for things that are overlooked with others
10	It makes me angry when I am treated worse than others

Table A2*Observer Sensitivity Items*

Nr.	Item
1	It bothers me when someone gets something they don't deserve
2	I am upset when someone does not get a reward he/she has earned
3	I cannot easily bear it when someone unilaterally profits from others
4	It takes me a long time to forget when someone else has to fix others' carelessness
5	It disturbs me when someone receives fewer opportunities to develop his/her skills than others
6	I am upset when someone is undeservingly worse off than others
7	It worries me when someone has to work hard for things that come easily to others
8	I ruminate for a long time when someone is treated nicer than others for no reason
9	It gets me down to see someone criticized for things that are overlooked with others
10	I am upset when someone is treated worse than others

Table A3*Beneficiary Sensitivity Items*

Nr.	Item
1	It disturbs me when I receive what others ought to have
2	I have a bad conscience when I receive a reward that someone else has earned
3	I cannot easily bear it to unilaterally profit from others
4	It takes me a long time to forget when others have to fix my carelessness
5	It disturbs me when I receive more opportunities than others to develop my skills
6	I feel guilty when I am better off than others for no reason
7	It bothers me when things come easily to me that others have to work hard for
8	I ruminate for a long time about being treated nicer than others for no reason
9	It bothers me when someone tolerates things with me that other people are being criticized for
10	I feel guilty when I receive better treatment than others

Table A4*Perpetrator Sensitivity Items*

Nr.	Item
1	It gets me down when I take something from someone else that I don't deserve
2	I have a bad conscience when I deny someone the acknowledgment he or she deserves
3	I cannot stand the feeling of exploiting someone
4	It takes me a long time to forget when I allow myself to be careless at the expense of someone else
5	It disturbs me when I take away from someone else the possibility of developing his or her potential
6	I feel guilty when I enrich myself at the cost of others
7	It bothers me when I use tricks to achieve something while others have to struggle for it
8	I ruminate for a long time when I treat someone less friendly than others without a reason
9	I have a bad conscience when I criticize someone for things I tolerate in others
10	I feel guilty when I treat someone worse than others

Manuscript B

Buchholz, M., Magraw-Mickelson, Z., Stolz, C., Süssenbach, P., & Gollwitzer, M. (submitted).

Taking back control: Findings on the cognitive, behavioral, and motivational consequences of victim sensitivity. Manuscript submitted for publication in *Social Justice Research*.

Taking back control: Findings on the cognitive, behavioral, and motivational consequences of victim sensitivity

Merle Buchholz¹, Zoe Magraw-Mickelson², Christopher Stolz¹, Philipp Süßenbach³, and
Mario Gollwitzer²

¹Department of Psychology, Philipps-Universität Marburg, Germany

²Department of Psychology, Ludwig-Maximilians-Universität München, Germany

³Department for Human Resources/Health/Social Sciences, Fachhochschule des Mittelstands
(University of Applied Sciences) Bielefeld, Germany

Author Note

Correspondence concerning this article should be addressed to Merle Buchholz,
Department of Psychology, Philipps-Universität Marburg, Gutenbergstraße 18, 35032
Marburg, Germany. Email: merle.buchholz@uni-marburg.de

Acknowledgements: We would like to thank Madita Schneider (Study 1) and Ineke
Postert (Study 2) for assistance in data collection.

Funding: This research was supported by two grants from the German Research
Foundation to Mario Gollwitzer and Philipp Süßenbach (grants no. GO 1674/5-1 and SU
902/2-1) as well as by the Research Training Group “Breaking Expectations” awarded to
Philipps-Universität Marburg by the German Research Foundation (project number:
290878970-GRK2271, projects 4 and 6).

Conflict of interest: We have no conflicts of interest to disclose.

Abstract

Being duped is an aversive experience which people are motivated to avoid. For this reason, especially people with a high fear of exploitation (i.e., people high in victim sensitivity; VS) often behave pre-emptively selfish and defensive in socially uncertain situations. Because the cognitive and motivational processes underlying such defensiveness have received little attention so far, we conducted two studies aiming to close this research gap. In Study 1 ($n_1 = 84$), we used virtual reality technology to examine whether social distancing, hostile interpretations of an interaction partner's intentions and behavior, and legitimizing cognitions regarding own selfish reactions (as elements of a suspicious mindset) mediate the effect of VS on uncooperativeness. Results did not show the expected mediation, but VS was still related to hostile information processing and fear of exploitation. In Study 2 ($n_2 = 273$), we extended these findings by showing that defensive reactions of people high in VS can be attenuated if a sense of control is reinforced. Together, the two studies crucially expand our knowledge of the defensive motivational system in victim-sensitive individuals.

[176 words]

Keywords: fear of exploitation, victim sensitivity, suspicious mindset, virtual reality, need for control

Taking back control: Findings on the cognitive, behavioral, and motivational consequences of victim sensitivity

It is not possible to go through life without encountering any injustice, inequality, or victimization. In interpersonal transactions, possibilities for exploitation are manifold, and the sad truth is that one can never be too sure of other people's motives. Whether one falls for a phony bargain, becomes a victim of fraud, is cheated on by a romantic partner, or needs to do the lion's share of work in a group task because the other team members refrain from contributing: being duped means being the "sucker". Thus, people are well advised to not trust others blindly, and to withdraw their trust and cooperation if there is reason to believe that an interaction partner might be untrustworthy.

The perception of having been taken advantage of is certainly painful and highly aversive (Vohs et al., 2007). Not only does being "suckered" result in a disadvantageous outcome for the victimized person, but it also implies that one has been a fool for not having seen it coming. For this reason, the feeling of being duped is often accompanied by self-blame and other self-conscious emotions like shame, embarrassment, and regret. Consequently, people are strongly motivated to avoid feeling duped, and when they anticipate or expect exploitation, they often behave in a pre-emptively defensive way.

For example, such pre-emptive defensiveness or pre-emptive selfishness has been discussed as one explanation for social loafing, which describes the phenomenon that people tend to decrease effort when working in groups compared to when working alone (Latané et al., 1979). One reason why people loaf in groups appears to be that they expect others to do the same: they work less hard because they are afraid that their group members would otherwise "free ride" on their efforts. Stated differently, the feeling of being duped is so aversive that one would rather slack off and put up with a worse outcome than being the "sucker" who carries the rest of the group (Kerr, 1983).

Importantly, there are stable inter-individual differences in the motivation to avoid exploitation. While some people hate the idea of falling prey to other people's malicious intentions, others simply care less about potentially being duped. The personality trait capturing such a latent fear of exploitation is called "victim sensitivity" (VS). In line with the above reasoning, VS has been shown to predict uncooperative and even immoral behavior in socially uncertain interdependence situations, when the risk of being duped is (considered to be) high. For instance, Gerlach et al. (2012) found victim-sensitive people to be less forgiving and more revengeful following a relational transgression committed by a romantic partner or close friend. Similarly, Faccenda and colleagues (2009) showed that after being fouled, soccer players high in victim sensitivity are more inclined to engage in unsportsmanlike, transgressive behaviors themselves. Additionally, victim-sensitive individuals also tend to make more egoistic choices in social dilemma games and other enticing situations (Fetchenhauer & Huang, 2004; Gollwitzer & Rothmund, 2011; Gollwitzer et al., 2005, Gollwitzer et al., 2009; Maltese et al., 2016; Rothmund et al., 2011), and more likely engage in counterproductive work behavior when feeling treated unfairly in organizational contexts (Lavelle et al., 2018).

A theoretical explanation for these findings is provided by the Sensitivity to Mean Intentions (SeMI) Model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). According to this model, victim-sensitive individuals become more easily suspicious, and in turn often behave selfishly and uncooperatively. More specifically, people high in VS are assumed to be particularly sensitive towards contextual cues indicating untrustworthiness. When such cues are present, a suspicious mindset (consisting of an attributional bias regarding others' malevolence, legitimizing cognitions regarding one's own selfish behaviors, and an avoidance-related motivational state) is activated in victim-sensitive people, and in this state of suspiciousness, there is less adherence to social norms and fairness principles. Thus, the

suspicious mindset is theoretically conceptualized as the mechanism explaining the effect of VS on uncooperative behavior. Put differently, the SeMI model describes a moderated mediation: when cues of untrustworthiness are present (“moderator”), VS activates a suspicious mindset (“mediator”), which then results in the withdrawal of trust and cooperation.

Empirical Evidence for the Suspicious Mindset

In situations that require trust, victim-sensitive individuals’ strong fear of exploitation leads them to react defensively and hostile in order to protect themselves from possible victimization. In other words: victim-sensitive people do everything to avoid being duped. However, whereas a number of studies have looked at the “total effect” of VS (x untrustworthiness cues) on uncooperative behavior, the precise cognitive and motivational processes underlying such defensiveness have received less attention. As a consequence, evidence for the suspicious mindset and its three components (i.e., attributional bias, legitimizing cognitions, avoidance motivation) has been so far indirect.

The few studies that investigated the suspicious mindset have mostly used scenario-based approaches. For example, Gerlach et al. (2012, Study 3) applied the SeMI model to research on interpersonal forgiveness and instructed their participants to imagine that a close friend had harmed them (for instance by disclosing a secret). After imagining this scenario, participants were asked how they interpreted the friend’s intentions, that is, whether they thought the friend’s post-transgression behavior was indicative of ulterior motives (attributional bias), and whether or not unforgiving reactions on their behalf would be justified (legitimization). In line with the SeMI model’s predictions, victim sensitivity was negatively related to participants’ willingness to forgive, and this relationship was (partially) mediated by hostile interpretations and self-protective legitimizing cognitions.

Following these findings, Maltese et al. (2016; Study 1) used ambiguous scenarios to show that heightened expectancies of injustice can (in part) account for victim-sensitive people's uncooperative behavior in trust games as well. As expected by the authors, participants high in VS were more inclined to endorse and anticipate unjust outcomes when reading the scenarios (at least when a suspicious mindset had been activated before), and this tendency to form hostile interpretations mediated the effect of VS on trust and cooperation.

Lastly, Rothmund et al. (2011) were able to demonstrate that experiencing virtual aggression from a victim's perspective resulted in reduced trust expectations as well. In their study, participants first played a video game sequence in which they were either betrayed by a video game character or not. Afterwards, they took part in a common goods dilemma game in the real world. Here, participants had to indicate (1) how much they wanted to contribute, and (2) how much they expected other players to invest in this situation. Again, a significant indirect effect was found: when confronted with virtual untrustworthiness, participants were more reluctant to cooperate in the subsequent dilemma game, and this was mediated by heightened mistrust expectancies with regard to others' investments.

In sum, previous research has attempted to elucidate the suspicious mindset and its components. Although these studies have shed some light on the cognitive dynamics of people high in VS, they have focused mainly on one element (namely the attributional bias), oftentimes by using scenario-based approaches with low ecological validity. Therefore, more research is needed in this context, which looks for stronger and more direct support regarding the three elements of the mindset. This is what Study 1 of the present research aims to do. In Study 2, we go a step further by illuminating the motivational processes operating behind the self-protective concerns of victim-sensitive people.

Motivational Processes Underlying the Pre-Emptive Defensiveness

What exactly motivates people to avoid exploitation and to react pre-emptively selfish when the risk for being duped is considered to be high? One of the few considerations of this question is given by Vohs et al. (2007). As briefly discussed above, these authors argue that people are motivated to avoid exploitation because the feeling of being duped is an aversive emotional state related to self-blame and other negative self-conscious emotions. More specifically, Vohs et al. (2007) posit that feeling duped arises from the perception of having been taken advantage of in a social context after one *chose to trust* the other person. Therefore, the victim is partly to blame for its situation: the unfavorable outcome could have been avoided if a different decision had been made (i.e., not trusting the other party), and for this reason, the experience of victimization results in self-recrimination. Stated differently, being deceived by another person has threatening implications for the self; it suggests that one is socially incompetent or stupid. Because people are reluctant to adopt such negative self-views, they will hence be motivated to avoid being the “sucker”. Thus, Vohs et al. (2007) postulate that individuals with a high fear of exploitation (i.e., people high in VS or “sugrophobia”) are defensive because the anticipation of being exploited is highly aversive and threatens their positive self-image.

In contrast, we argue that victim-sensitive people’s defensiveness is not so much about avoiding self-image threats, but more about maintaining a *sense of control*. As a vast amount of research shows, people generally prefer to have (perceived) control over themselves, their environment, and the outcome of social situations (e.g., Fiske, 2003; Kay et al., 2008; Rothbaum et al., 1982; Skinner, 1995, 1996). Victim-sensitive individuals, however, might be particularly motivated to be in control: As Schmitt, Neumann, and Montada (1995) demonstrated, a higher sensitivity to own unjust disadvantages was associated with a higher need for control. In our opinion, socially uncertain interdependence situations are particularly threatening to this need for control. In these situations, untrustworthiness cues act as a

warning signal for vulnerability and victimization—an aversive outcome imposed by others. The individual should hence feel at the mercy of untrustworthy forces, and, as a consequence, should be motivated to (re)establish a sense of control by preventing the expected exploitation (c.f. Rothbaum et al., 1982; Skinner, 1996). Thus, we think that the defensiveness shown by people high in VS reflects a means of coping with the threat posed by lowered levels of perceived control.

Study 2 was specifically designed to test these two competing hypotheses (i.e., anticipated exploitation threatens victim-sensitive people’s positive self-image vs. their need for control) directly against each other. Notably, if we can explain why people high in VS tend to react with self-protective and vengeful tendencies in socially uncertain situations, then we might be able to attenuate them. Understanding the motivation behind victim-sensitive individuals’ defensiveness becomes especially relevant when considering that a high fear of exploitation crucially impairs social interactions.

The Present Research

The aim of the present research is to further illuminate the defensiveness mechanism underlying victim sensitivity. More specifically, we pursue two goals: in Study 1, we simultaneously investigate all three elements of the suspicious mindset by using immersive virtual reality technology. A virtual environment has the advantage that it allows to study participants’ behavior in an ecologically realistic way while maintaining a high degree of experimental control (McCall & Blascovich, 2009); thus, Study 1 goes far beyond what previous research on the suspicious mindset has achieved. In addition, we introduce social distancing as a behavioral measure of the mostly neglected third component of the mindset, avoidance motivation. Accordingly, two hypotheses are being tested in this study: first, we expect to replicate the established effect of VS on trust and cooperation. Second, we

hypothesize that this relation is mediated by hostile interpretations, legitimizing cognitions, and participants' approach-avoidance behavior in a virtual world.

In Study 2, we go a step further by exploring why people high in VS actually show such pre-emptive defensiveness. As discussed above, two hypotheses are plausible in this context: on the one hand, it may be that victim-sensitive individuals are defensive because they are afraid that being duped threatens their self-image (c.f. Vohs et al., 2007). On the other hand, persons high in VS might be defensive because they want to maintain a sense of control over what is happening in their social environment.

Study 1

Study 1 was designed to investigate the suspicious mindset in a more immersive and direct way than has been done in past research. More specifically, the focus of Study 1 was on participants' avoidance motivation; a component of the suspicious mindset that has been neglected in previous studies. As a behavioral measure for this avoidance-related motivational state, social distancing was used. We expected that in a state of suspiciousness, individuals high in VS would (physically) distance themselves from others because in trust-relevant situations, any interaction partner represents a source of potential exploitation (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013).

To activate such a state of suspiciousness in our participants, we confronted them with untrustworthiness cues in the beginning (see below). Then, they had to solve easy tasks in a fully immersive virtual world together with an ostensible partner while we tracked the distance between both players. The use of an immersive virtual environment allowed us to accurately, continuously, and unobtrusively measure participants' distancing behavior (c.f. Bailenson et al., 2003; Kane et al., 2012), and, at the same time, to standardize the behavior of the ostensible partner (all of the confederate's movements were pre-programmed). After completing the tasks in the virtual world, participants played a modified trust game (Berg et

al., 1995). Victim sensitivity and other personality traits were measured several days or weeks prior to the lab appointment in a separate online questionnaire.

In sum, Study 1 tested the mediation described in the SeMI model (Gollwitzer et al., 2013). Thus, we predicted that after being confronted with untrustworthiness cues, people high in VS would keep more distance towards their partner's avatar (M1), would interpret the other person's intentions and behavior in a more distrustful, hostile way (M2), and would stronger endorse legitimizing cognitions regarding their own uncooperative behavior (M3). We further assumed that these mechanisms would explain (i.e., mediate) the negative effect of VS on participants' cooperativeness in the trust game (DV).

Participants

Taking into account the efforts that come with the use of the VR lab, our goal was to recruit as many participants as feasible within our six month timeframe (November 2018 to May 2019). The study that was ostensibly about "coordination in a virtual world" was advertised via university mailing lists, flyers, and an advertisement in a local newspaper. The study consisted of two parts: an online questionnaire (which was filled out by 269 participants), and a subsequent lab appointment several days or weeks later (here, 92 subjects took part). Participants whose datasets could not be unequivocally matched between those two parts were excluded from the sample. In addition, three participants had to be excluded because they were suspicious with regard to the confederate ($n=2$) or because of technical issues in the virtual reality lab ($n=1$). Thus, the final analysis sample consisted of 84 participants (48% male, 52% female). The majority of them (93%) were undergraduate students of a wide variety of disciplines; ages ranged from 18 to 50 years ($M=23.6$, $SD=4.6$ years). Participants received monetary compensation for their time, the exact amount (4€-10€) was dependent upon their decisions in the trust game.

Materials and Measures

As mentioned above, the study consisted of an online questionnaire in which personality traits were measured, and a lab appointment in which we assessed social distancing, hostile interpretations, legitimizing cognitions, and trusting behavior. Data from these different parts was connected via a personalized code that participants created on their own and that ensured (pseudo-)anonymity. For privacy reasons, this code was later deleted and replaced by a numerical code in the dataset.

Personality Traits

At the beginning of the online questionnaire, participants were informed that the study consisted of two parts and that it would take approximately one hour to complete the study in total. After giving informed consent, participants provided some demographic information and then filled out several trait measures, including victim sensitivity (VS), observer sensitivity (OS), general trust, and the Big Five. Victim and observer sensitivity were assessed with 10 items each (Justice Sensitivity Inventory; Schmitt et al., 2010), which were answered on a Likert scale from 0 (“totally disagree”) to 5 (“totally agree”). Example items are: “It bothers me when others receive something that ought to be mine” (VS) and “I am upset when someone does not get a reward that he/she has earned” (OS). The scales have good internal consistency (in the present study: Cronbach’s $\alpha=.84$ and $.89$, respectively). General trust ($\alpha=.81$) was measured with a German version of the General Trust Scale by Yamagishi and Yamagishi (1994) and was answered on a six-point Likert scale as well. Finally, neuroticism ($\alpha=.80$), extraversion ($\alpha=.85$), and openness for experience ($\alpha=.58$) were measured with the BFI-10 (Rammstedt & John, 2007); here, response scales ranged from 1 to 5. Several other scales were assessed as well but because they are not relevant for the present research, they will not be referred to further. At the end, participants had the possibility to sign up for the second part of the study, which was scheduled several days or weeks later. In addition, they were informed that the lab appointment would take place with a second participant.

Virtual Reality

In the lab, participants were informed that their partner, in reality a female confederate, would be a couple of minutes late but that they would start preparations without her. The confederate being unpunctual had two reasons: first, because it should raise doubts about her dependability and trustworthiness, and second, because participants should not see the other person in the “real” world (to avoid any positive or negative first impressions). For this reason, participants were also told not to talk to the other person once she arrived.

Next, participants received an information sheet, which informed them that they would first solve different coordination tasks in a virtual world and that they would play a (trust) game with their partner afterwards. They were also given a brief summary of the trust game instructions to make the possibility of a later exploitation salient. After participants gave informed consent, they put on a head mounted display (HMD) and were led to their starting position for the virtual tasks (while wearing the HMD, participants were unable to see any part of the “real” world). Next, the experimenter secretly informed the confederate via SMS or WhatsApp, who arrived at the lab approximately two minutes later. The confederate signed the informed consent as well and then (ostensibly) put on a second HMD¹.

Before starting the virtual environment, the experimenter explained that both players would be randomly assigned to an avatar (i.e., a human representation in the virtual world) but that they would only see the avatar of their counterpart. In fact, participants did not see any virtual representation of their own body. The confederate, however, was always represented by a female avatar with a neutral facial expression (see Figure 1). A neutral appearance was chosen because, as previous research has demonstrated, victim-sensitive

¹ We used two fully functional HMDs for the study (nVisor ST50 and HTC Vive); however, the second HMD was only a requisite and not connected to the virtual world. Thus, while participants were led to believe that the other player controlled the avatar they were seeing, the confederate was not even able to see the virtual world. Instead, our research assistants had been trained to imitate the pre-programmed avatar (to make it more realistic, the avatar followed recorded movements of a real person). As a result, participants saw the avatar moving in the virtual world and, at the same time, heard the confederate’s footsteps in the real world.

individuals perceive neutral faces as less trustworthy (Gollwitzer et al., 2012; Study 1). A neutral facial expression, although being a subtle cue, should thus activate a suspicious mindset in individuals high in VS, especially if combined with other untrustworthiness cues (like the tardiness of the confederate and the possibility of later exploitation in the trust game). Next, the experimenter started the virtual environment. Here, participants saw a spacious plaza with a fountain, plants, and some shops and restaurants (see Figure 1). In addition, a blue and a red cross on the ground indicated the starting points for the respective players (participant and confederate were facing each other when standing on their starting positions). We did not present participants with auditory input in the virtual world because they had to listen to the experimenter's instructions.

After familiarization with the virtual world in a warm-up task, participants' approach-avoidance behavior was assessed in three proxemic tasks. Here, participants were asked to (1) walk around their partner to read a combination of numbers and letters from their back (this task was validated in previous research, c.f. Bailenson et al., 2003), (2) to avoid additional agents (i.e., virtual humans controlled by a computer algorithm) that walked through the virtual place, and (3) to approach their partner up to the point where the closeness still felt natural and comfortable for them (c.f. Sommer, 2002; Uzzell & Horne, 2006).

Importantly, the two players took turns in performing these tasks; thus, only participant or avatar were moving at any given time (in Tasks 1 and 3 the participant went first, in Task 2 the confederate). In all proxemic tasks, the position and orientation of participant and avatar were continuously tracked and this data was used to compute the minimum distance for each task². We chose minimum distance as a measure for social distancing because it has been successfully used in both traditional proxemic research as well as in studies investigating interpersonal distance in virtual environments (c.f. Bailenson et al.,

² Participants' position was tracked from the front right of the HMD. We only used length and width coordinates for distance computations; thus, height differences between participants did not influence distance measures.

2001; Bailenson et al., 2003).

Trust Game

After the tasks in the virtual world, participants were escorted to another room in which they were seated in front of a computer screen. They were informed that they would now play a game with their partner and that the points they won in this game would be added to their 2€ compensation fee: for every 10 points they gained, they would be paid 1€ extra. The experimenter then left the room, ostensibly to tend to the other person who still waited in the virtual reality lab³, while participants familiarized themselves with the instructions of the game. The rules were explained as follows: both players would make two decisions, one in the role of “sender” (i.e., trustor) and one in the role of “receiver” (i.e., trustee), but they would not be informed about the decisions of their partner until the end of the game. As sender, they would receive 20 points, which they could either keep (those points would be added to their payment) or send to the receiver. More specifically, participants could send 0, 5, 10, 15, or 20 points that were then quadrupled by the experimenter. The receiver, on the other hand, could then decide to either keep all or to share any amount of the quadrupled points with the sender.

After participants read these instructions, they first made their decision as sender (this decision functioned as our DV) and then as receiver. However, because they were not informed about the decisions of their alleged partner during the game, participants had to make the second decision for all possible outcomes (strategy method). After each decision, participants answered a short follow-up survey about their feelings and thoughts during the respective decision (e.g., “I am satisfied with my decision” and “I was afraid of being taken advantage of by the other player”; see Table A1).

SeMI Components

³ Participants always left the virtual reality lab first (while the confederate still wore the HMD) to make sure that participant and confederate did not interact with each other outside of the experimental tasks.

After the trust game, participants answered a series of questions about their perception of the other person and her behavior in the virtual world. Four of these items assessed hostile interpretations (i.e., “Do you think the other person is trustworthy?” (reversed), “Do you think the other person has hostile motives?”; Cronbachs $\alpha = .75$), two items measured legitimizing cognitions (i.e., “Do you think it is justified not wanting to work with the other person?”, “Suppose you were to work on a project with the other person (e.g., for a seminar). Do you find it justified to refuse?”; Cronbachs $\alpha = .67$), the rest (e.g., “I had the impression that the other person kept some distance to me in the virtual world”) was included for exploratory purposes. All of these follow-up items (see also Appendix A; Tables A2 and A3) were answered on six-point Likert scales. Finally, participants were probed for suspicion, fully debriefed, and received their compensation fee⁴.

Results and Discussion

To test our hypotheses, we correlated victim sensitivity (and other relevant personality traits) with our three mediator variables and our DV in a first step. These correlations, as well as means and standard deviations of the respective variables, can be found in Table 1. Notably, the intercorrelations between personality measures resembled earlier findings (c.f. Schmitt et al., 2005): For example, victim sensitivity was positively related to observer sensitivity and neuroticism but uncorrelated with extraversion or openness.

As can be seen in Table 1, minimum distance⁵, i.e., the shortest distance that participants maintained between themselves and the avatar of the other person, was uncorrelated with victim sensitivity ($r = .11, p = .31$). Interestingly, though, exploratory

⁴ The calculation of the fee was standardized but depended on the decisions in the trust game. Thus, participants received between 4€ and 10€.

⁵ Looking at this first potential mediator, we found, rather unexpectedly, that the minimum distances measured in the different proxemic tasks were uncorrelated with each other ($r = -.04$ to $.19$, all $p_s > .08$). For this reason, we did not compute a mean across tasks as originally planned; instead, we focused on Task 1 (where participants had to read the number from the other person’s back) because this task has been used and validated in previous research on interpersonal distance using virtual reality technology.

analyses of the follow-up items revealed that whereas participants high in VS showed no avoidance tendencies themselves, they still perceived *the other person* to be more distant and reserved. Thus, despite the fact that avatar movements were pre-programmed, victim-sensitive participants more strongly had the impression that the other person kept some distance to them in the virtual world ($r = .25, p = .02$). Unexpectedly, we also found observer sensitivity and openness for experience to be positively correlated with the social distance measure, which implies that participants high in OS and openness kept more distance to the avatar than participants low in these traits. However, the openness scale showed a very low internal consistency in the present study ($\alpha = .58$); therefore, correlations with openness should be interpreted with caution.

With regard to the other two suspicious mindset components, we found that VS was positively correlated with hostile perceptions of the other player ($r = .24, p = .03$) but uncorrelated with legitimizing cognitions ($r = .14, p = .21$). Thus, participants high in VS interpreted the other person's intentions and behavior in a more negative, mistrustful way but they did not consider uncooperativeness toward the other person as more justifiable. However, this may have not been necessary because participants high in VS showed no uncooperative behavior in our study anyway: contrary to our hypothesis and to the results of previous research, we found no correlation between victim sensitivity and trust in the trust game ($r = .09, p = .40$). In fact, the decision to send points was unrelated to any of the personality traits measured, as well as to the suspicious mindset variables (see Table 1).

Mediation Analysis

Although the bivariate correlations between our variables of interest did not always show the expected pattern, we conducted a multiple mediation analysis using the PROCESS macro (Hayes, 2018) in a second step. To test whether the hypothesized negative relation between VS and trust was mediated by hostile interpretations of the other person's behavior,

legitimizing cognitions of one's own uncooperative reactions, and social distancing tendencies, we included all three mediator variables simultaneously into the model while controlling for observer sensitivity, general trust, neuroticism, and extraversion⁶. An overview of the results can be found in Figure 2. Overall, only 10% of the variance in trust was explained by our model. Contrary to our hypothesis, victim sensitivity had no (total or direct) effect on trust. With regard to the mediators, we found VS to predict a heightened readiness to infer mean intentions ($\beta = .27, p = .04$), but no significant effects on legitimizing cognitions or social distancing were found ($p = .14$ and $p = .78$, respectively). In accordance with the correlation analyses, none of these mediator variables significantly influenced the amount of points sent in the trust game (although the positive effect of social distancing reached marginal significance). For this reason, it is not surprising that no indirect effect of VS on trust could be demonstrated: Bootstrapping results showed that the 95% confidence intervals always contained zero. Thus, we were unable to replicate the established negative relationship between victim sensitivity and trust, and we also did not find support for the assumption that social distancing, hostile interpretations, and legitimizing cognitions (as elements of a suspicious mindset) mediated this hypothesized relation.

In sum, victim sensitivity was not (directly or indirectly) related to distrust in our findings. However, when taking a closer look at our DV it became clear that a ceiling effect (which reduces variation in trusting behavior) might have contributed to this lack of findings: while more than 50% of participants decided to send all of their 20 points to their partner, only one person sent nothing⁷. Importantly, exploratory analyses of the follow-up items still

⁶ Openness for experience was not included as a covariate because of the scale's low internal consistency.

⁷ The median was 20, the mean was 16.25 ($SD=4.962$), and the skewness index was -1.13 ($SE=0.263$), suggesting a high negative skewness and a significant deviation from the normal distribution (Shapiro-Wilks' $W=0.75, p<.001$). For this reason, we used the Box-Cox transformation (Box & Cox, 1964) as described in Study 2. This procedure estimates a parameter (lambda or "Box-Cox parameter") which yields a quasi-optimal fit of a variable against a normal distribution. We used Free Statistics Software (v1.2.1) to estimate this parameter (Wessa, 2016), which equaled $\lambda=1.47$ in our case. The R code and the Box-Cox Normality Plot for our data is stored here: <https://www.freeststatistics.org/blog/index.php?v=date/2020/Oct/12/t1602496919kua7mdhjy2olhr5.htm/>.

revealed effects on a cognitive and emotional level: with higher VS, participants reported to be less satisfied with their decision (“I am satisfied with my decision”; $r = -.29, p = .01$), and to have greater fear of exploitation in the trust game context (“I was afraid of being taken advantage of by the other player”; $r = .24, p = .03$). In other words, participants high in VS were afraid that the other person might take advantage of them, but they did not act on this fear. One reason for this could be that the untrustworthiness cues that we provided participants with were not quite strong enough to trigger self-protective behavior. Thus, it is possible that uncooperative behavior, unlike hostile information processing, has a higher threshold of suspiciousness. Another explanation is given by Dunning et al. (2014): the authors argue that humans sometimes show trust in others (especially strangers) because they believe that norms of respect force them to do so. Thus, in many situations, individuals do not really expect benevolence from a cooperation partner but feel obliged to show trusting behavior nonetheless. It is therefore possible that victim-sensitive participants in our study sent points to their partner because this was what they felt they *should* do, although it was not what they really *wanted* to do.

Although we were unable to find support for our main hypotheses, the results of Study 1 still provided some important insights into the defensiveness mechanism underlying victim sensitivity. For instance, our results show that while VS did not predict social distancing, participants high in VS perceived the confederate to be physically distant—despite the fact that the partner’s behavior was standardized. Together with the finding that victim sensitivity promoted hostile interpretations of the other person’s behavior and intentions in general, this lends further support for the “inaccuracy hypothesis” (Gollwitzer et al., 2012; Gollwitzer et al., 2013): According to this notion, VS implies an asymmetrical sensitivity toward

Importantly, results were virtually identical when using the transformed trust variable. To simplify interpretation, we therefore report results with the DV in its original metric.

untrustworthiness cues, which results in less accurate predictions of other people's cooperativeness (i.e., a bias in social judgments). In addition, it is also noteworthy that victim-sensitive participants showed a heightened readiness to infer mean intentions although we confronted them with only slight cues of untrustworthiness.

In conclusion, we were able to demonstrate that victim sensitivity was related to more unfavorable perceptions of the other player as well as to fear of exploitation, but we did not find any effects on a behavioral level (i.e., participants high in VS showed no social distancing and no withdrawal of trust and cooperation). Hence, while these results still suggest that victim-sensitive individuals' information processing is characterized by a pronounced suspiciousness in socially uncertain situations, no conclusions about the underlying motivational processes could be drawn. Therefore, Study 2 was conducted to investigate why individuals high in VS are more likely than others to act pre-emptively defensive and how their self-protective strikes may be attenuated.

Study 2

Study 2 was designed to test the hypothesis that restoring a sense of control over what happens in one's social environment can effectively alleviate victim-sensitive individuals' tendency to distrust others. Participants in this study played a trust game with an unknown fellow participant. Victim sensitivity was measured as in Study 1. Since previous research (e.g., Gollwitzer & Rothmund, 2011; Rothmund et al., 2011; Süssenbach & Gollwitzer, 2015) suggests that victim sensitivity predicts distrust only when a suspicious mindset has been activated, half of the participants were confronted with cues of untrustworthiness (i.e., a respective facial expression and a message; see below). We wanted to activate a strong suspicious mindset in order to test our hypothesis (i.e., that restoring a sense of control alleviates suspiciousness and distrust among victim-sensitive individuals) as strictly and conservatively as possible.

To restore a sense of control, we used an established control-affirmation procedure adapted from Kay, Gaucher, Napier, Callan, and Laurin (2008): here, participants are prompted to recall a situation in which they had recently experienced a sense of control (see Methods section below). Although this procedure has been successfully used in prior research, we figured that recalling a situation in which participants had control might not only affirm their momentary sense of control, but may also and inadvertently boost their self-regard. For instance, recalling a situation in which one has shown moral courage and protected another person from being bullied may effectively reinstall a sense of control, but may also (re)affirm one's moral virtues and self-regard (e.g., Monin & Miller, 2001). Therefore, in order to test whether restoring a sense of control is *specifically* responsible for alleviating the effect of suspiciousness on distrust among victim-sensitive individuals, we also implemented an experimental condition in which participants were self-affirmed (using a procedure adapted from Monin et al., 2008).

Thus, Study 2 used a 2 (suspicious mindset, no suspiciousness) \times 3 (control-affirmation, self-affirmation, no-affirmation) between-subjects design with victim sensitivity as a continuous moderator variable. We predicted that, when a suspicious mindset has been activated, victim-sensitive individuals would be less trustful towards a fellow participant when no affirmation occurred, and that this effect was alleviated specifically after restoring participants' sense of control. Thus, we hypothesized a suspiciousness \times affirmation \times victim sensitivity three-way interaction effect.

Participants

Undergraduate students from a mid-size German university were invited via email to participate in an online study about "social interactions, self-views, and emotions". In exchange for participation, five online book vouchers worth 50€ each were raffled among all participants. The invitation contained a link to the survey, which was active for two months

(May to July 2014). Three-hundred and twenty-one students clicked on this link; 291 finished the study. Participants who needed less than 1 minute ($n=12$) or more than 2 hours to complete the online survey ($n=5$) were discarded from further analyses. One additional case was deleted because this participant entered an invalid number (i.e., 12) in the trust game (values could only range between 0 and 10, see below). Thus, the final sample consisted of 273 participants (173 female; 100 male). Ages ranged between 18 and 64 years ($M=25$, $SD=7.1$ years). Students were enrolled in a broad range of undergraduate programs, including psychology ($n=18$). We aimed at recruiting as many participants as possible within the two-month time frame. Assuming a small- to medium-size three-way interaction effect (i.e., $f^2=.03$), as in Gollwitzer & Rothmund (2011; Study 2), with a power of 80% would have required a sample size of 315 if $\alpha=5\%$ (calculated with G*Power 3.1; Faul et al., 2009). To ensure a sufficient power to detect the hypothesized effect with a sample size of 273, we increased our significance level to $\alpha=10\%$; here, the power is 83%.

Materials and Measures

Trust Game

After participants gave their informed consent, they were introduced to the rules of the trust game. They were told that they would be paired with a second (female⁸) participant from another university in the south of Germany, with which we would be cooperating on a research project. Participants were informed that they would play two rounds with their partner: in Round 1, participants would be the “sender” (i.e. trustor) and their partners would be “receivers” (i.e., the trustee). In Round 2, roles would be switched. Unbeknownst to participants, there was no second round, and there also was no other participant. Participants received 10 lottery tickets, of which they could send any number (between 0 and 10) to their

⁸ Partners were female because the majority of participants in our study was female. Our rationale for not varying the partner’s sex was that doing so would have reduced the statistical power for detecting the suspiciousness \times affirmation \times victim sensitivity three-way interaction effect.

partner. The number of tickets sent would be tripled by the experimenter so that receivers would now have their own 10 tickets plus the tripled number of received tickets from their partner (i.e., the real participant). At that point, receivers would have to decide whether or not they want to split the total number of tickets equally between both players (this version of the trust game has been described by Kuwabara, 2005; for an application in research on the SeMI model, see Gollwitzer & Rothmund, 2011; Study 2). Participants were also told that the number of lottery tickets they had in the end would determine their chances of winning the 50€ book voucher: the more tickets, the higher the likelihood of winning one of these vouchers.

Suspiciousness Manipulation

Next, participants saw a picture that their respective partner has allegedly taken of herself during their experimental session. These pictures were selected from the Amsterdam Dynamic Facial Expression Set (ADFES; van der Schalk et al., 2011), a stimulus set containing 648 emotional (facial) expressions displayed by 22 models (for a validation, see Wingenbach et al., 2016). For the present purpose, two pictures were selected (see Figure 3), one displaying a neutral facial expression (left panel) and the other displaying joy (right panel). More precisely, the neutral facial expression was used to induce a “suspicious mindset” among participants, whereas the joyful facial expression was used to prevent the induction of suspiciousness. Using similar stimuli, Gollwitzer et al. (2012; Study 1) showed that a neutral facial expression is indeed sufficient to induce a suspicious mindset among victim-sensitive individuals.

To make our manipulation even stronger, participants received a (fake) message from their partner, which she allegedly wrote to introduce herself. These messages were designed to induce a suspicious mindset (“*Hi, my name is Heike. I’m 25 and studying business administration in Tübingen. It’s awesome that I have the chance to win something here.*”) vs.

to reduce the likelihood of suspicions being raised (“*Hi, my name is Sarah. I’m 25 and studying social work in Tübingen. It’s awesome that we both have the chance to win something here.*”). Prior research showed that a message in which one’s partner focuses on maximizing their own profit (as in the suspicious mindset condition) vs. on maximizing the team’s collective profit (as in the no suspiciousness condition) can effectively amplify the effect of distrust on uncooperative responses (Parks et al., 1996).

Affirmation Manipulation

Next, participants were randomly assigned to either a control-affirmation, a self-affirmation, or a no-affirmation (control) condition. Participants in the *control-affirmation condition* were instructed to recall a positive event from the last couple months for which they were personally responsible, in which they experienced a “sense of control over what had happened” (adapted from Kay et al., 2008). Participants in the *self-affirmation condition* were prompted to recall a positive event from the last couple months in which they were able to display a virtue or express a value that was personally relevant for them, such as a sense of humor, social competence, musical talent, etc. (taken from Monin et al., 2008). Participants in the *no-affirmation (control) condition* were asked to list all the things they do on a typical weekday. In all three conditions, participants were asked to use the free-response field provided to write a short text (100 words max). These texts were screened and coded for validity and credibility. All participants followed the instructions and wrote a principally credible text, which was in line with the respective prompt.

After writing these texts, participants were re-familiarized with the rules of the trust game. Then, they actually played the trust game and decided how many lottery tickets they would want to transfer to their partner. This measure was our dependent variable.

Personality Traits

Finally, participants completed a battery of personality trait measures, including the

10-item Victim Sensitivity Scale ($\alpha=.83$), the 10-item Observer Sensitivity Scale ($\alpha=.84$), both taken from Schmitt et al. (2010), and the 6-item General Trust Scale ($\alpha=.82$) taken from Yamagishi and Yamagishi (1994). Response scales on all measured variables ranged between 1 (“totally disagree”) to 6 (“totally agree”). After responding to a number of control questions (e.g., “What do you think this study was about?”), participants were informed about their final amount of lottery tickets, thanked, and debriefed.

Results and Discussion

Looking at our central DV, participants’ decisions in the trust game, many of them ($n=119$, i.e., 44%) decided to transfer all of their lottery tickets (i.e., 10) to their partner. Only 7 participants (2.6%) decided to keep their 10 tickets for themselves. The median was 8, the mean was 7.57 ($SD=2.7$), and the skewness index was -0.83 ($SE=0.15$), suggesting a moderate to high negative skewness and a significant deviation from the normal distribution (Shapiro-Wilks’ $W=0.832$, $p<.01$). Therefore, we used a procedure proposed by Box and Cox (1964) to find the optimal transformation for our variable,⁹ which, in our case, was

$$Y_{\text{trans}} = \frac{(Y + 0.5)^{\lambda} - 1}{\lambda}$$

with Y denoting the trust measure in its original metric, and Y_{trans} denoting the transformed variable (see also Fox & Weisberg, 2011; Velilla, 1993). Means, standard deviations, and correlations between the transformed trust variable and the three personality traits measured are displayed in Table 2.

As expected, victim sensitivity was positively related to observer sensitivity and negatively related to general trust, which is why we used observer sensitivity and general trust as covariates in the regression model reported below. In addition, the transformed trust variable

⁹ This procedure estimates a parameter (lambda or “Box-Cox parameter”) which yields a quasi-optimal fit of a variable against a normal distribution. We used Free Statistics Software (v1.2.1) to estimate this parameter (Wessa, 2016), which equaled $\lambda=1.38$ in our case. The R code and the Box-Cox Normality Plot for our data is stored here: <https://www.freestatistics.org/blog/index.php?v=date/2020/May/13/t1589377509uogl22wdbqyqvfm/>.

was negatively related to victim sensitivity and positively related to general trust across all experimental conditions.

We predicted that when a suspicious mindset has been activated, victim-sensitive individuals would be less trustful towards a fellow participant when no affirmation occurred, and that this effect would be alleviated specifically by a control-affirmation (vs. a self-affirmation) procedure. To test the hypothesized suspiciousness \times affirmation \times victim sensitivity three-way interaction effect, we contrast-coded the suspiciousness conditions (-1=no suspiciousness, 1=suspicious mindset) as well as the affirmation conditions so that Contrast 1 reflected an effect of control-affirmation vs. baseline (control-affirmation=1, self-affirmation=0, no-affirmation=-1) and Contrast 2 reflected an effect of self-affirmation vs. baseline (control-affirmation=0, self-affirmation=1, no-affirmation=-1). Victim sensitivity was standardized to facilitate the interpretation of regression weights. Observer sensitivity and general trust were included as covariates. The estimated regression coefficients are reported in Table 3. These results suggest that only a control-affirmation procedure (but not a self-affirmation procedure) was able to make victim-sensitive individuals just as trustful as their victim-insensitive counterparts, as indicated by the significant VS \times Suspiciousness \times Contrast_1 interaction effect. This effect is also graphically displayed in Figure 4. Simple slope analyses further corroborated this interpretation: while victim sensitivity was negatively related to trust in the suspicious mindset condition when no affirmation occurred ($B=-1.834$, $p=.015$), this effect disappeared in the control affirmation condition ($B=-0.459$, $p=.596$). Thus, although participants high in VS tend to distrust others in socially uncertain situations, this pre-emptive defensiveness can actually be attenuated if a sense of control is (re-)established.

General Discussion

Being duped is a highly aversive, painful experience which people are motivated to avoid. For this reason, especially people with a high fear of exploitation (i.e., people high in

victim sensitivity) tend to act pre-emptively selfish and hostile in situations in which the risk for being “suckered” is considered to be high. However, although past research has repeatedly investigated the relation between VS and uncooperativeness in socially uncertain situations (e.g., Faccenda et al., 2009; Fetchenhauer & Huang, 2004; Gollwitzer et al., 2005; Lavelle et al., 2018), the assumed mechanism (i.e., the suspicious mindset) has received far less attention (but c.f. Gerlach et al., 2012; Maltese et al., 2016, Rothmund et al., 2011). To close this research gap, we conducted two studies investigating the cognitive and motivational processes underlying such pre-emptive defensiveness. Together, the findings crucially expand our knowledge of the defensive motivational system in victim-sensitive individuals.

The Suspicious Mindset

To better understand how victim sensitivity translates into selfish and uncooperative behavior, we examined possible mediating processes in Study 1. More precisely, we used virtual reality technology to create an immersive environment in which participants interacted with another person to unobtrusively measure social distancing (as a form of self-protective, avoidant behavior). Although our main focus was on this approach-avoidance behavior, we also assessed hostile interpretations of the other person’s behavior and intentions, and legitimizing cognitions with regard to own uncooperative reactions in self-report (i.e., the other two components that are assumed to constitute the suspicious mindset). Finally, participants played a trust game with their alleged partner over real money.

Notably, the results did not support our hypotheses for the most part. First, we were unable to show the expected correlation between victim sensitivity and social distancing. More specifically, VS only predicted hostile and mistrustful perceptions of the other person, but not legitimizing cognitions or avoidance behavior. However, although participants high in VS showed no distancing tendencies themselves, they still perceived *the other person* to keep their distance, which was by design a purely subjective impression. Thus, our findings

corroborate the notion that VS promotes a bias in social judgments: victim-sensitive individuals focus strongly on untrustworthiness cues, which distorts their perceptions of their social environment (Gollwitzer et al., 2012; Gollwitzer et al., 2013).

Second, trust rates were overall very high but unrelated to any of the measured variables. In contrast to previous studies, victim-sensitive participants did not send less points to their partner in the trust game, even though they reported more fear of exploitation than victim-insensitive participants. Thus, we were unable to replicate the established effect of victim sensitivity on distrust, and, as a consequence, we were also unable to show that the suspicious mindset mediated this relation. However, there are a number of possible reasons that might explain why the expected direct and indirect effects of VS on trust were absent in our findings. For instance, the sample size may have simply been too small to ensure adequate statistical power for the detection of the hypothesized effects. Monte Carlo simulations suggest that several hundred participants are required if small to medium sized effects are tested in a (multiple) mediation model (Fritz & MacKinnon, 2007; Ma & Zeng, 2014; Thoemmes et al., 2010). For this reason, future studies should aim at recruiting a much larger sample than the one used in the present study.

In addition, the SeMI model as well as previous research (c.f. Gollwitzer & Rothmund, 2011; Rothmund et al., 2011; Süssenbach & Gollwitzer, 2015) suggest that VS predicts uncooperative and distrustful behavior only if there are sufficient indications that an interaction partner might be untrustworthy. The untrustworthiness cues that participants encountered in Study 1, though, were relatively subtle. Although similar stimuli have been successfully used in previous studies (Gollwitzer et al., 2012), it may be that our manipulations failed to raise strong enough suspicions in victim-sensitive participants. This could be due to other factors in the design that might have mitigated the untrustworthy cues: For example, participants interacted with the other person for quite some time before playing

the trust game, and, as previous research demonstrates, people are more likely to trust a specific counterpart if they are not anonymous to each other (Glaeser et al., 2000).

Additionally, although the “partner” arrived late (cuing untrustworthiness), the confederate followed all the experimenter’s instructions and worked with the participant to solve the tasks in the virtual world (cuing trustworthiness). To eliminate the possibility that distrustful, uncooperative behavior simply has a higher threshold of untrustworthiness cues than in our design, future studies should aim at activating an unambiguous suspicious mindset.

Nonetheless, the fact that victim-sensitive participants reported more fear of exploitation and less satisfaction with their own decision in the trust game corroborates the notion that our manipulation was (at least, in part) successful: participants high in VS *were* more fearful and suspicious of their partner’s intentions, although they did not act on it.

In conclusion, victim sensitivity was exclusively related to an attributional bias regarding others’ malevolence in Study 1, which confirms and complements past findings on the suspicious mindset in victim-sensitive individuals (Gerlach et al., 2012; Maltese et al., 2016; Rothmund et al., 2011). However, although VS predicted more unfavorable perceptions of the other player, we were not able to demonstrate that this bias actually mediated behavioral trust and cooperation. One reason for this lack of findings may be low statistical power. Another explanation could be that defensive behavior simply has a higher threshold of untrustworthiness cues: while hostile information processing may be triggered even in situations containing only slight meanness cues (like in the present study), more evidence for mean intentions might be needed to actually justify the withdrawal of trust and cooperation. Therefore, future research should rigorously test the conceptualization of a suspicious mindset in highly-powered designs containing strong and unambiguous cues of untrustworthiness.

Motivational Basis of Defensiveness

To examine why exactly victim-sensitive people tend to react defensively in socially

uncertain situations and how such pre-emptively selfish behavior may be alleviated, a second study was conducted. More precisely, Study 2 was designed to test whether victim-sensitive individuals are defensive because (1) they are afraid that being exploited threatens their self-image (c.f. Vohs et al., 2007) or because (2) they want to maintain a sense of control over their social environment. To compare these two possible motivations directly against each other, we employed an experimental design in which participants were either self-affirmed or control-affirmed (or not affirmed at all) after the activation of a suspicious mindset (vs. after no induction of suspiciousness), and then instructed to play a trust game with an ostensible partner. Results demonstrated, as expected, that when a sense of control was reinforced, victim-sensitive individuals' tendency to distrust others was effectively alleviated. The affirmation of a positive self-image, on the other hand, did not make participants high in VS more trustful in the presence of untrustworthiness cues. Thus, these findings corroborate the notion that victim-sensitive people are motivated to react in vengeful and self-protective ways because anticipated exploitation threatens their need for control (and not their self-image). In this context, it has to be noted that we activated a strong suspicious mindset in Study 2 in order to test our hypotheses as strictly and as conservatively as possible. The fact that affirming a sense of control still made victim-sensitive individuals as trustful as their victim-insensitive counterparts thus suggests that this is an effective way to mitigate their elevated suspiciousness. Thus, it may be possible to develop interventions capable to modify and reduce victim-sensitive people's distrust tendencies in the long-run. We consider this an important avenue for future research.

Limitations

The present research further corroborates the SeMI model's predictions and expands our knowledge about the pre-emptive defensiveness shown by victim-sensitive individuals, but of course there are also some limitations. First and foremost, the sample size of Study 1

was rather small. For this reason, the suspicious mindset and its components should be investigated in more highly powered studies in the future, especially if multiple mediation models are to be applied. This will allow to draw conclusive inferences about the underlying defensiveness mechanism.

Another constraint lies in the fact that the minimum distance measures were uncorrelated across proxemic tasks in Study 1, although all of these tasks were designed to assess participants' distancing behavior. However, in fact they differed in their demands: While Task 1 (reading the number from the other person's back) and Task 3 (moving toward the other person and stopping at a comfortable distance) specifically asked participants to approach their partner, Task 2 focused on the avoidance of additional virtual humans (therefore increasing the possibility that participants were too distracted to pay attention to the confederate or to their physical proximity). In addition, due to the not randomized order of the tasks, it is possible that throughout the tasks, participants grew accustomed to approaching the other person—and to getting approached by her—which could have influenced their approach-avoidance behavior in later tasks. Some evidence that supports this reasoning is the finding that the mean minimum distance was 0.46 meters in Task 1 and 0.45 meters in Task 3, although previous research has shown that people typically approach more closely to the back (Task 1) than to the front (Task 3) of other people (Bailenson et al., 2003). Thus, we focused on the minimum distance assessed in the first proxemic task, which should represent the most valid measure for social distancing in our study¹⁰. Nevertheless, future research may want to use more than one measure of approach-avoidance behavior (or even additional operationalizations of avoidance motivation like the avoidance of eye contact).

Conclusion

¹⁰ Results vary slightly as a function of the distance measure used. For example, the minimum distances assessed in Tasks 2 and 3 were uncorrelated with Observer Sensitivity and Openness. In addition, the direct effect of social distancing on trust in the mediation analysis was marginally significant (although unexpectedly positive) for the minimum distance of Task 1, but far from significant for the distances of Tasks 2 and 3.

A large body of empirical findings shows that a high fear of exploitation (i.e., high victim sensitivity) results in uncooperative and hostile behaviors in socially uncertain situations. Importantly, this pre-emptive defensiveness implies a focus on others and their behavior: instead of worrying about themselves and the impression they make, people high in VS are more concerned about what others are up to. In other words, defensiveness is an inherently social phenomenon, not a merely self-centered phenomenon. Investigating this social sensitivity is important not only from a theoretical but also from an applied perspective: if we understand which cognitive and motivational processes are at work in victim-sensitive individuals, we may be able to develop effective interventions capable of preventing or attenuating adverse effects of VS on trust and cooperation. The present studies provide important insights in this context. Most importantly, they show that boosting a sense of control helps victim-sensitive people to overcome their habitual suspiciousness. Stated differently, providing people with a sense of control might prove to be an effective tool to improve social interactions and to break vicious cycles of non-cooperation.

Declarations

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence, 10*(6), 583-598. <https://doi.org/10.1162/105474601753272844>
- Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin, 29*(7), 819-833. <https://doi.org/10.1177/0146167203029007002>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. <https://doi.org/10.1006/game.1995.1027>
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B, 26*(2), 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122-141. <http://doi.org/10.1037/a0036673>
- Faccenda, L., Pantaléon, N., & Reynes, E. (2009). Significant predictors of soccer players' moral functioning from components of contextual injustice, sensitivity to injustice and moral atmosphere. *Social Justice Research, 22*, 399-415. <https://doi.org/10.1007/s11211-009-0105-4>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences, 36*(5), 1015–1029. [https://doi.org/10.1016/S0191-8869\(03\)00197-1](https://doi.org/10.1016/S0191-8869(03)00197-1)

- Fiske, S. T. (2003). Five core social motives, plus or minus five. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated social perception: The Ontario Symposium* (Vol. 9, pp. 233-246). Erlbaum.
- Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (2nd ed.). Sage.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233-239. <https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Gerlach, T. M., Allemand, M., Agroskin, D., & Denissen, J. J. A. (2012). Justice sensitivity and forgiveness in close interpersonal relationships: The mediating role of mistrustful, legitimizing, and pro-relationship cognitions. *Journal of Personality, 80*(5), 1373-1413. <https://doi.org/10.1111/j.1467-6494.2012.00762.x>
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics, 115*(3), 811-846. <https://doi.org/10.1162/003355300554926>
- Gollwitzer, M., & Rothmund, T. (2009). When the need to trust results in unethical behavior: The Sensitivity to Mean Intentions (SeMI) model. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 135–152). Information Age Publishing.
- Gollwitzer, M., & Rothmund, T. (2011). What exactly are victim-sensitive persons sensitive to? *Journal of Research in Personality, 45*(5), 448–455. <https://doi.org/10.1016/j.jrp.2011.05.003>
- Gollwitzer, M., Rothmund, T., Alt, B., & Jekel, M. (2012). Victim sensitivity and the accuracy of social judgments. *Personality and Social Psychology Bulletin, 38*(8), 975–984. <https://doi.org/10.1177/0146167212440887>

- Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality, 43*(6), 999–1005. <https://doi.org/10.1016/j.jrp.2009.07.003>
- Gollwitzer, M., Rothmund, T., & Süssenbach, P. (2013). The Sensitivity to Mean Intentions (SeMI) model: Basic assumptions, recent findings, and potential avenues for future research. *Social and Personality Psychology Compass, 7*(7), 415–426. <https://doi.org/10.1111/spc3.12041>
- Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical effects of justice sensitivity perspectives on prosocial and antisocial behavior. *Social Justice Research, 18*(2), 183–201. <https://doi.org/10.1007/s11211-005-7368-1>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). The Guilford Press.
- Kane, H. S., McCall, C., Collins, N. L., & Blascovich, J. (2012). Mere presence is not enough: Responsive support in a virtual world. *Journal of Experimental Social Psychology, 48*(1), 37-44. <https://doi.org/10.1016/j.jesp.2011.07.001>
- Kay, A. C., Gaucher, D., Napier, J. L., Callan, M. J., & Laurin, K. (2008). God and the government: Testing a compensatory control mechanism for the support of external systems. *Journal of Personality and Social Psychology, 95*(1), 18-35. <https://doi.org/10.1037/0022-3514.95.1.18>
- Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology, 45*(4), 819–828. <https://doi.org/10.1037/0022-3514.45.4.819>
- Kuwabara, K. (2005). Nothing to fear but fear itself: Fear of fear, fear of greed and gender effects in two-person asymmetric social dilemmas. *Social Forces, 84*(2), 1257-1272. <https://doi.org/10.1353/sof.2006.0020>

- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37(6), 822–832. <https://doi.org/10.1037/0022-3514.37.6.822>
- Lavelle, J. J., Harris, C. M., Rupp, D. E., Herda, D. N., Young, R. F., Hargrove, M. B., Thornton-Lugo, M. A., & McMahan, G. C. (2018). Multifoci effects of injustice on counterproductive work behaviors and the moderating roles of symbolization and victim sensitivity. *Journal of Organizational Behavior*, 39(8), 1022-1039. <https://doi.org/10.1002/job.2280>
- Ma, Z.-w., & Zeng, W.-n. (2014). A multiple mediator model: Power analysis based on Monte Carlo simulation. *American Journal of Applied Psychology*, 3(3), 72-79. <https://doi.org/10.11648/j.ajap.20140303.15>
- Maltese, S., Baumert, A., Schmitt, M. J., & MacLeod, C. (2016). How victim sensitivity leads to uncooperative behavior via expectancies of injustice. *Frontiers in Psychology*, 6, 2059. <https://doi.org/10.3389/fpsyg.2015.02059>
- McCall, C., & Blascovich, J. (2009). How, when, and why to use digital experimental virtual environments to study social behavior. *Social and Personality Psychology Compass*, 3(5), 744-758. <https://doi.org/10.1111/j.1751-9004.2009.00195.x>
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33-43. <https://doi.org/10.1037/0022-3514.81.1.33>
- Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology*, 95(1), 76-93. <https://doi.org/10.1037/0022-3514.95.1.76>
- Parks, C. D., Henager, R. F., & Scamahorn, S. D. (1996). Trust and reactions to messages of intent in social dilemmas. *Journal of Conflict Resolution*, 40(1), 133-150.

<https://doi.org/10.1177/0022002796040001007>

- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Rothbaum, F., Weisz, J. R., & Snyder, S. S. (1982). Changing the world and changing the self: A two-process model of perceived control. *Journal of Personality and Social Psychology, 42*(1), 5-37. <https://doi.org/10.1037/0022-3514.42.1.5>
- Rothmund, T., Gollwitzer, M., & Klimmt, C. (2011). Of virtual victims and victimized virtues: Differential effects of experienced aggression in video games on social cooperation. *Personality and Social Psychology Bulletin, 37*(1), 107-119. <https://doi.org/10.1177/0146167210391103>
- Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*, 211–238. <https://doi.org/10.1007/s11211-010-0115-2>
- Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment, 21*(3), 202-211. <https://doi.org/10.1027/1015-5759.21.3.202>
- Schmitt, M. J., Neumann, R., & Montada, L. (1995). Dispositional sensitivity to befallen injustice. *Social Justice Research, 8*(4), 385-407. <https://doi.org/10.1007/BF02334713>
- Skinner, E. A. (1995). *Perceived control, motivation, and coping*. Sage Publications.
- Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology, 71*(3), 549-570. <https://doi.org/10.1037/0022-3514.71.3.549>
- Sommer, R. (2002). Personal space in a digital age. In R. B. Bechtel & A. Churchman (Eds.), *Handbook of environmental psychology* (pp. 647-660). John Wiley & Sons, Inc.

- Süssenbach, P., & Gollwitzer, M. (2015). Us(ed): The role of victim sensitivity in potentially exploitative intergroup relationships. *Group Processes & Intergroup Relations*, *18*(2), 241-255. <https://doi.org/10.1177/1368430214556700>
- Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, *17*(3), 510-534. <https://doi.org/10.1080/10705511.2010.489379>
- Uzzell, D., & Horne, N. (2006). The influence of biological sex, sexuality and gender role on interpersonal distance. *British Journal of Social Psychology*, *45*(3), 579-597. <https://doi.org/10.1348/014466605X58384>
- van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expressions Set (ADFES). *Emotion*, *11*(4), 907-920. <https://doi.org/10.1037/a0023853>
- Velilla, S. (1993). A note on the multivariate Box-Cox transformation to normality. *Statistics and Probability Letters*, *17*, 259-263. [https://doi.org/10.1016/0167-7152\(93\)90200-3](https://doi.org/10.1016/0167-7152(93)90200-3)
- Vohs, K. D., Baumeister, R. F., & Chin, J. (2007). Feeling duped: Emotional, motivational, and cognitive aspects of being exploited by others. *Review of General Psychology*, *11*(2), 127-141. <https://doi.org/10.1037/1089-2680.11.2.127>
- Wessa, P. (2016). *Box-Cox Normality Plot (v1.1.12) in Free Statistics Software (v1.2.1)*. [Software]. http://www.wessa.net/rwasp_boxcoxnorm.wasp/
- Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV): A set of videos expressing low, intermediate, and high intensity emotions. *PLoS ONE*, *11*(1), e0147112. <https://doi.org/10.1371/journal.pone.0147112>
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129–166. <https://doi.org/10.1007/BF02249397>



Fig.1 Virtual environment and female avatar used in Study 1. Both the plaza and the avatar were taken from built-in packages of the Vizard virtual reality software (WorldViz). Printed by permission.

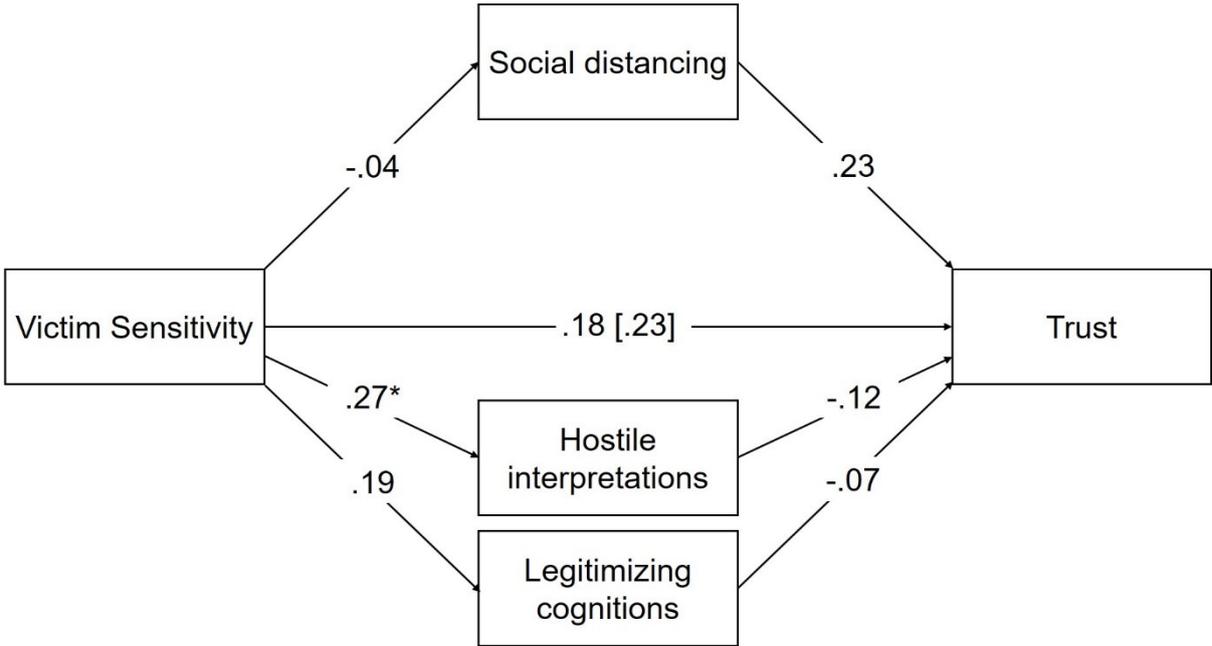


Fig.2 Multiple mediation of the effect of VS on trust via social distancing, hostile interpretations, and legitimizing cognitions. Standardized regression coefficients are depicted. $*p < .05$ (two-tailed).



Fig.3 Stimulus material used in Study 2 to induce (left panel) vs. not induce (right panel) suspiciousness. Stimuli were selected from the Amsterdam Dynamic Facial Expression Set (ADFES; van der Schalk et al., 2011). Reprinted by permission.

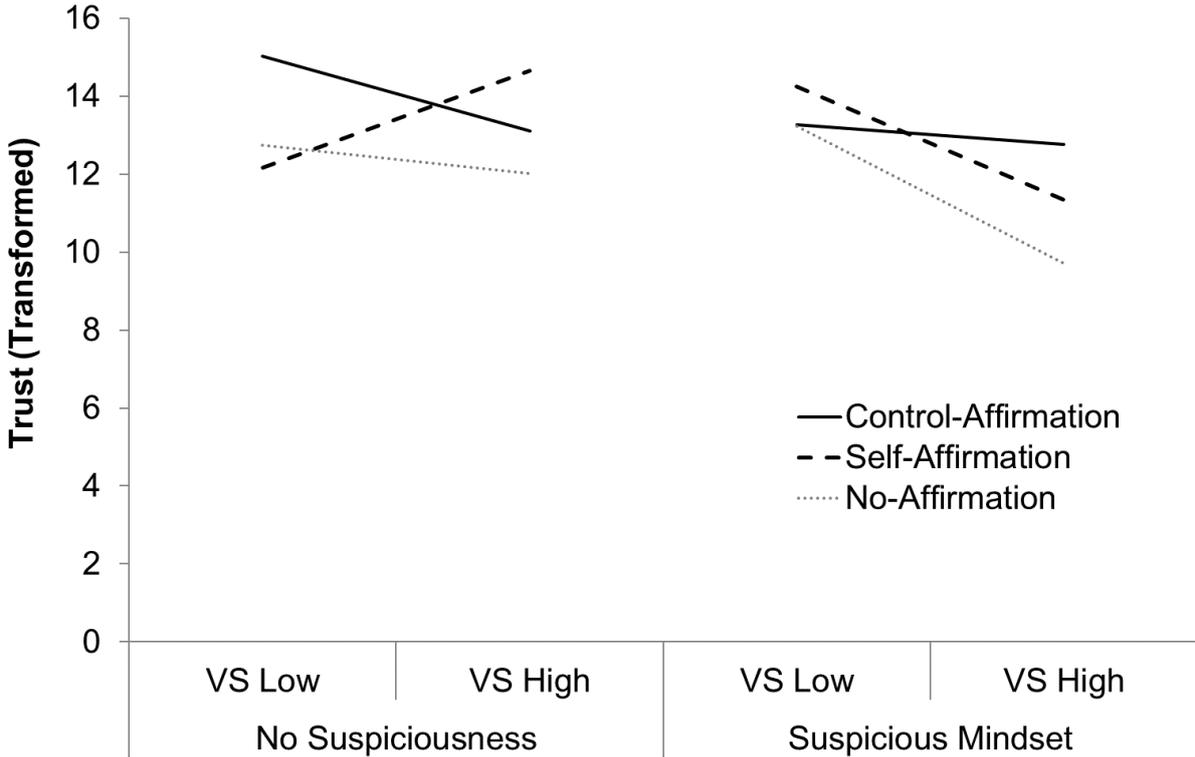


Fig.4 Predicted means visualizing the suspiciousness × affirmation interaction effect for participants low vs. high in victim sensitivity (± 1 *SD* around the sample mean; Study 2).

Table 1*Means, Standard Deviations, and Correlations Between Measured Variables (Study 1)*

Variable	<i>M</i> (<i>SD</i>)	Correlations										
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
(1) Victim sensitivity	2.92 (0.83)	1										
(2) Observer sensitivity	2.96 (0.90)	.41**	1									
(3) General trust	2.96 (0.82)	-.18	.08	1								
(4) Neuroticism	3.17 (1.06)	.45**	.20	-.09	1							
(5) Extraversion	2.97 (1.10)	-.04	.14	.27*	-.38**	1						
(6) Openness	3.82 (0.94)	.04	.07	.14	.02	.13	1					
(7) Min. distance	0.46 (0.14)	.11	.27*	.07	.13	.10	.24*	1				
(8) Hostile interpretation	2.14 (0.66)	.24*	-.15	-.34**	.10	-.03	.15	-.08	1			
(9) Legitimizing cognitions	2.24 (1.25)	.14	-.25*	-.34**	.09	-.07	-.05	-.03	.37**	1		
(10) Trust	16.25 (4.96)	.09	-.02	.004	-.01	-.08	-.04	.19	-.09	-.06	1	

Notes. $N=84$. Response scales on personality scales ranged from 0-5 (victim sensitivity, observer sensitivity, general trust) or from 1-5 (Big-5). Min. distance refers to the shortest distance participants kept to the avatar when reading the number on her back (in meters). Hostile interpretation was measured with 4 items; legitimizing cognitions with two items (on a scale from 1-6). Trust refers to the sender decision in the trust game (measured on a scale from 0-20 points). * $p<.05$; ** $p<.01$ (two-tailed).

Table 2*Means, Standard Deviations, and Correlations Between Measured Variables (Study 2)*

Variable	<i>M (SD)</i>	Correlations		
		(2)	(3)	(4)
(1) Trust (transformed)	12.63 (5.58)	-.14*	.10	.18**
(2) Victim sensitivity	3.59 (0.85)	1	.22**	-.17**
(3) Observer sensitivity	3.90 (0.86)	.22**	1	.19**
(4) General trust	3.90 (0.85)	-.17**	.19**	1

Notes. $N=273$. Response scales on personality scales ranged from 1-6.

* $p < .05$; ** $p < .01$ (two-tailed).

Table 3*Results from the Moderated Regression Model (Study 2)*

Predictor	<i>B</i>	<i>SE(B)</i>	<i>t</i>
Constant Term	12.852	0.339	37.956**
Victim Sensitivity (VS)	-0.592	0.358	-1.654*
Suspiciousness Manipulation	-0.428	0.339	-1.263
Observer Sensitivity	0.516	0.349	1.477
General Trust	0.812	0.345	2.358**
Contrast_1 (Control-Affirmation vs. Baseline)	0.681	0.498	1.366
Contrast_2 (Self-Affirmation vs. Baseline)	0.247	0.475	0.519
VS × Suspiciousness	-0.561	0.343	-1.637
VS × Contrast_1	-0.015	0.501	-0.030
VS × Contrast_2	0.488	0.495	0.986
Suspiciousness × Contrast_1	-0.096	0.502	-0.190
Suspiciousness × Contrast_2	0.120	0.476	0.252
VS × Suspiciousness × Contrast_1	0.913	0.505	1.810*
VS × Suspiciousness × Contrast_2	-0.781	0.495	-1.577

Notes. $N=273$. * $p<.10$ ** $p<.05$. Victim Sensitivity, Observer Sensitivity, and General Trust are z -standardized.

Appendix A**Table A1***Exploratory Follow-Up Items Trust Game Decisions*

Item
I am satisfied with my decision.
I was afraid of being taken advantage of by the other player.
I could not resist the temptation to keep as much money as possible for myself.
I think my counterpart would have acted exactly like me.

Table A2*Exploratory Follow-Up Items Behavior in the Virtual World*

Item
I had the impression that the other person kept some distance to me in the virtual world.
The other person seemed to be aware of where I was in the virtual world.
The other person did not want to get too close to me in the virtual world.
The other person acted in the virtual world the way I expected them to.
It seemed as if the other person was executing the tasks as well as they could.
In the virtual world, the other person did not seem to care where I was.
The distance the other person kept in the virtual world influenced my behavior.
The other person did not bother to work with me to solve the tasks in the virtual world.
The other person came too close to me in the virtual world.
The behavior of the other person had no influence on my behavior in the virtual world.

Table A3*Items SeMI Components*

Item
Would you like to get to know the other person better? ⁺
How likeable is the other person? ⁺
How well would you get along with the other person? ⁺
Do you think one can trust the other person? (reversed) *
Do you think the other person has hostile motives? *
Do you think the other person is trustworthy? (reversed) *
Suppose you were to work on a project with the other person (e.g., for a seminar). Do you find it justified to refuse? [†]
Do you think it is justified not wanting to work with the other person? [†]
Do you think the other person is a difficult cooperation partner? *

*These items assessed hostile interpretations

[†]These items measured legitimizing cognitions

⁺These items were filler items

Manuscript C

Buchholz, M., Bergmann, N., Schubö, A., & Gollwitzer, M. (submitted). Victim sensitivity predicts attention allocation towards violations of untrustworthiness expectancies.

Manuscript submitted for publication in *Social Justice Research*.

**Victim Sensitivity Predicts Attention Allocation Towards Violations of
Untrustworthiness Expectancies**

Merle Buchholz¹, Nils Bergmann¹, Anna Schubö¹, and Mario Gollwitzer²

¹Department of Psychology, Philipps-Universität Marburg, Germany

²Department of Psychology, Ludwig-Maximilians-Universität München, Germany

Author Note

Some of the findings reported here were presented at the 17th Conference of the Social Psychology Section (FGSP 2019) in Cologne, Germany. Correspondence concerning this article should be addressed to Merle Buchholz, Department of Psychology, Philipps-Universität Marburg, Gutenbergstraße 18, 35032 Marburg, Germany. Email: merle.buchholz@uni-marburg.de

Funding: This research was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, project number 290878970-GRK 2271, Projects 4 and 9) to Philipps-Universität Marburg.

Availability of data and material: The data underlying all findings, as well as codebooks and research materials are publicly available on the Open Science Framework at <http://doi.org/10.17605/OSF.IO/BY5P3>.

Conflict of interest: We have no conflicts of interest to disclose.

Abstract

Victim Sensitivity (VS)—a personality trait reflecting the anxious expectation of being exploited by other people—predicts social distrust and uncooperativeness, but may also reflect a strong latent motivation to trust others. Therefore, information implying a violation of untrustworthiness expectations (i.e., trust-related attributes being associated with an untrustworthy-looking face) may be more motivationally relevant for victim-sensitive persons than information implying a violation of trustworthiness expectations (i.e., distrust-related attributes being associated with a trustworthy-looking face). To test this hypothesis, participants' ($n=69$) eye movements were recorded while they saw trustworthy or untrustworthy facial expressions and words that either confirmed or violated the expectation elicited by the respective face. Results show that victim sensitivity was associated with an attentional bias towards information violating untrustworthiness expectations, but not with an attentional bias towards information violating trustworthiness expectations. The study provides first evidence that victim sensitivity influences how trustworthiness-related social information is differentially processed.

[149 words]

Keywords: victim sensitivity, social information, untrustworthiness expectations, expectancy violation, attention allocation

Victim Sensitivity Predicts Attention Allocation Towards Violations of Untrustworthiness Expectancies

People differ in the extent to which they fear being exploited: some hate the idea of falling prey to other people's selfish intentions, while others simply do not care. The personality trait that captures such a latent fear of exploitation has been referred to as "victim sensitivity". While previous research has mainly looked at the behavioral consequences of being dispositionally victim-sensitive, the question how exactly victim-sensitive people process social information related to trustworthiness or untrustworthiness, and, more importantly, how they process social information that *violates* their (un)trustworthiness expectancies¹, has been largely neglected so far. The present study aims to close this research gap.

According to the Sensitivity to Mean Intentions (SeMI) Model, victim-sensitive people expect others to be malevolent and selfish, and they are strongly motivated to avoid being exploited (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). As a consequence, individuals high in victim sensitivity are assumed to react more sensitively than persons low in victim sensitivity towards contextual cues that indicate untrustworthiness. In other words, the model proposes that victim-sensitive people become more easily suspicious, which ultimately makes them behave less cooperatively towards others.

A growing number of findings is in line with these assumptions (e.g., Fetchenhauer & Huang, 2004; Gollwitzer & Rothmund, 2011; Gollwitzer et al., 2005, 2009). For instance, victim-sensitive persons are less forgiving after transgressions in close relationships, reflecting a differential tendency to infer ulterior motives (Gerlach et al., 2012). In addition, people high in victim sensitivity behave uncooperatively even when confronted with only

¹ In this manuscript, the terms *expectancy* and *expectation* are used interchangeably. It should be noted, though, that expectations are more frequently defined as explicit, verbalized constructs, whereas expectancies are rather understood as something implicit that can be present without full awareness (Rief et al., 2015).

slight cues of untrustworthiness (Gollwitzer et al., 2009), and they rate neutral and angry looking faces to be less trustworthy compared to participants low in victim sensitivity (Gollwitzer et al., 2012; Study 1). This biased processing may explain why victim-sensitive people tend to underestimate other people's cooperativeness (Gollwitzer et al., 2012; Study 2). These and other findings (c.f. Maltese et al., 2016) suggest that victim sensitivity shapes information processing in a way that is personality-congruent (Rusting, 1998): victim-sensitive people are asymmetrically sensitive towards cues indicating selfishness and untrustworthiness, and they frequently adopt a "distrust mindset".

One of these cues is an interaction partner's facial expression: people swiftly draw inferences about a person's trustworthiness from their facial expression (Todorov et al., 2009; Willis & Todorov, 2006). However, the predictive validity of facial cues is far from perfect: even a grumpy-looking fellow can eventually turn out to be a nice person, and a truly selfish soul may be hidden under a nice and friendly appearance. So what if the judgment turns out to be wrong? How do victim-sensitive people process such expectancy violations?

Victim Sensitivity and Expectancy Violations

Assuming that victim-sensitive people are asymmetrically sensitive towards untrustworthiness (vs. trustworthiness) cues, it appears plausible to assume that victim-sensitive people also have a better memory for other people's untrustworthy (compared to trustworthy) behavior. However, previous research suggests an effect in the opposite direction. For instance, Süßenbach et al. (2016) showed that victim-sensitive individuals are particularly likely to remember information that *violates* their untrustworthiness expectations. In this study, participants saw pictures of male targets, and each target was accompanied by either a positive (i.e., trustworthiness-related, such as "scientist") or a negative (i.e., untrustworthiness-related, such as "trickster") social label. Approximately five seconds later, participants learned that the target had recently committed either a prosocial act (e.g.,

“rescued a kid that fell into a frozen pond”) or an antisocial act (e.g., “stole valuable items from the apartments of older people”). Across trials, the act (prosocial vs. antisocial) was uncorrelated with the label (i.e., trustworthy vs. untrustworthy). In a subsequent surprise memory test, individuals high (vs. low) in victim sensitivity were more likely to correctly remember targets with a negative social label who committed a prosocial act—a *positive expectancy violation*—and less likely to correctly remember negatively labeled targets committing an antisocial act. Memory for *negative expectancy violations*, that is, positively labeled targets committing an antisocial act, did not differ between participants high vs. low in victim sensitivity. This suggests that victim-sensitive individuals are especially receptive towards positive expectancy violations, but not towards negative expectancy violations.

In a second study, Süssenbach et al. (2016) expanded these findings: in a first phase, participants saw targets accompanied by either a negative (untrustworthiness-related) or a positive (trustworthiness-related) label just as in Study 1; in a subsequent phase, they received additional information about whether the target had committed a prosocial or an antisocial act. The targets’ trustworthiness was rated once after Phase 1 and a second time after Phase 2. Changes in trustworthiness ratings depended on participants’ victim sensitivity: participants high (vs. low) in victim sensitivity updated their trustworthiness perceptions more strongly after positive expectancy violations (i.e., negatively labeled targets committing prosocial acts) than after negative expectancy violations (i.e., positively labeled targets committing antisocial acts). In sum, these two studies suggest that people high in victim sensitivity do not focus solely on untrustworthiness-related cues; instead, violations of untrustworthiness expectations seem to have an even bigger impact on memory and impression updating.

Motivational Relevance of Expectancy Violations

The finding that victim sensitivity predicts source memory for positive, but not for negative expectancy violations, is at odds with a number of other studies. Results from trust

research, for example, indicate that humans identify and remember “cheaters”—people who violate social contracts—particularly well (e.g., Mealey et al., 1996; Oda, 1997). From an evolutionary perspective, it is indeed adaptive to pay more attention to potentially malevolent than to potentially benevolent interaction partners (Cosmides & Tooby, 1989). Hence, it seems surprising that victim-sensitive people show enhanced memory for prosocial behavior that violates untrustworthiness expectancies.

The solution for this apparent paradox may be deduced from an untested assumption of the SeMI Model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013). According to this model, victim sensitivity is rooted in (1) a *generalized expectation* that other people are untrustworthy and harbor mean intentions and, at the same time, (2) a *strong need* or motivation to trust others. Put differently, while victim-sensitive people expect others to be untrustworthy, they would love to live in a world in which other people can be trusted. Information contradicting their negative expectations about other people thus resonates with their strong need to trust; this type of expectancy violation is more *motivationally relevant* for victim-sensitive than for victim-insensitive people.

Importantly, Süßenbach et al. (2016) have focused only on source memory (Study 1) and impression updating (Study 2), but they never looked at earlier stages of information processing, such as attention allocation. Notably, attention allocation is influenced to a large extent by the motivational relevance ascribed to a stimulus (e.g., Summerfield & Egner, 2009). Therefore, a strict test of the assumption that positive expectancy violations are particularly motivationally relevant for victim-sensitive individuals would be to show that victim sensitivity predicts attention allocation to these positive expectancy violations. We will test this hypothesis in the present study.

Attention Allocation to Motivationally Relevant Stimuli

Our assumption that the motivational relevance of a stimulus predicts attention allocation to it and, thus, causes a better source memory for it, is in line with a large number of findings in the literature. Due to limited processing capacities of the human brain, only a subset of the available visual information can be processed in detail, and, therefore, attentional filters are needed to select and prioritize those stimuli that are most relevant for the organism's goals and needs (e.g., Lavie & Dalton, 2014). This process is referred to as selective visual attention.

Notably, selective visual attention determines to a great extent which information is processed and encoded into long-term memory (Chun et al., 2011; Chun & Turk-Browne, 2007). Attentional processing, on the other hand, is influenced by past experiences and (implicitly) acquired knowledge as well: extracted regularities from the visual environment, which are stored in memory, can be used for guiding visual search and the selection of stimuli, a process which is referred to as statistical learning (e.g., Bergmann et al., 2019; Goujon et al., 2015; Theeuwes, 2018). This interdependence between memory and visual attention makes it likely that a source memory advantage for positive expectancy violations (as in Süsserbach et al., 2016) is preceded by attention allocation towards these violations (for more studies investigating the relationship between attention and memory, see Aly & Turk-Browne, 2016; Cabeza et al., 2008; Heuer & Schubö, 2018; Wolfe et al., 2007).

Even more importantly, selective visual attention is influenced by motivation and motivational significance (e.g., Brosch & Van Bavel, 2012; Dietze & Knowles, 2016; Feldmann-Wüstefeld et al., 2016; Lang et al., 1997; Summerfield & Egner, 2009), and expectation violations constitute a particularly salient class of motivational stimuli (Proulx et al., 2017). However, motivational relevance does not seem to be a static or invariant construct; instead, the significance of stimuli is highly context dependent and reflects the flexible motivational state of the perceiver (Brosch & Van Bavel, 2012). For instance, DeWall

et al. (2009) induced a specific motivational state by thwarting participants' need for social belonging. As a consequence, visual attention was preferentially allocated to desired, goal-relevant stimuli, that is, to cues signaling social acceptance and affiliation. Thus, whenever needs or goals are threatened, attention is allocated to information that resonates with the respective need or to possibilities for satisfying the thwarted need. In addition, individual differences in the strength or accessibility of needs and motivational concerns influence the relevance ascribed to cues or information and affect the amount of attentional capture as well (Brosch & Van Bavel, 2012).

The Present Study

Based on our theorizing and the findings reviewed here, we expect to find a similar bias for positive expectancy violations in the allocation of visual attention in victim-sensitive individuals. As argued in the SeMI model (Gollwitzer & Rothmund, 2009; Gollwitzer et al., 2013), victim-sensitive persons (in contrast to victim-insensitive persons) are characterized by a particularly strong need to trust others. Thus, whenever a social cue elicits an untrustworthiness expectation (e.g., a grumpy face), information that is inconsistent with this cue (e.g., a positive social attribute) becomes particularly motivationally relevant for victim-sensitive individuals. More specifically, we argue that victim sensitivity predicts attention allocation to stimuli suggesting a violation of untrustworthiness expectations (i.e., positive expectancy violations).

With regard to negative expectancy violations, our predictions are less straightforward. Based on our reasoning and previous findings (e.g., Süssenbach et al., 2016), victim sensitivity should predict preferential attention allocation only towards positive expectancy violations, but not towards negative expectancy violations. However, past research has also shown that victim-sensitive persons focus more strongly on information signaling untrustworthiness when evaluating other people's behavior and intentions than victim-

insensitive persons (Gollwitzer et al., 2012). Put differently: due to their latent fear of exploitation, people high in victim sensitivity are particularly attracted to cues associated with untrustworthiness and selfishness. Therefore, one could argue that victim sensitivity also predicts attention allocation towards information indicating that a person is not as trustworthy as one would expect.

To test these hypotheses, an eye tracking study was conducted because saccadic eye movements are tightly interlinked with attentional selection processes; thus, they are well suited as proxies for attentional processes (Chun et al., 2011; Deubel & Schneider, 1996). Participants were seated in front of a computer screen and an eye tracker that recorded saccades and fixations. On this screen, pictures of male faces with positive (i.e., trustworthy) or negative (i.e., untrustworthy) facial expressions appeared and, after a short time interval, the faces were complemented by words related to trustworthiness or untrustworthiness. Matches and mismatches between the words and the respective facial expressions were used to operationalize confirmations and violations of (un)trustworthiness expectations, respectively, and to compare the allocation of attention (measured by the eye movement pattern) towards these different kinds of stimuli as a function of victim sensitivity.

Method

Open Data

In accordance with open science principles, primary data from this project can be downloaded from the Open Science Framework at <http://doi.org/10.17605/OSF.IO/BY5P3>.

Power Analysis

We determined sample size by calculating an a priori power analysis using G*Power 3.0.8 (Faul et al., 2007). Based on pilot study data, medium-sized correlations between victim sensitivity and eye tracking parameters seemed realistic; therefore we used an effect size

estimate of $\rho=.30$, with an alpha of .05 (one-tailed) and power of 0.8. The recommended minimum sample size given these parameters was 67.

Participants

A total of 70 undergraduate students from a German university were recruited via email advertisement and participated in the study for partial course credit or financial compensation. The dataset of one participant had to be excluded due to a lack of knowledge of the German language; therefore, the final analysis sample consisted of 69 students (52 female, 17 male). All participants had normal or corrected-to-normal vision and were between 18 and 32 years old ($M_{age}=21.9$, $SD=2.78$).

Apparatus

The experiment was conducted in a dimly lit and sound-attenuated room. The participants were seated in front of a 32-inch monitor (1920 × 1080 pixels, 120 Hz) and placed their heads on a chinrest, so that their eyes were aligned with the center of the screen in a distance of 100 cm. We recorded eye movements of the participants' right eye with an EyeLink 1000 Plus eye tracker (SR Research Ltd.), recording with 1000 Hz and calibrated with the 13-point calibration procedure. E-Prime Professional 2 (Psychology Software Tools) was used for stimulus presentation and response collection.

Materials

We selected 120 computer-generated pictures of frontal male faces from two databases freely available to researchers who conduct non-profit, academic research (Original Computer Generated Faces; see Oosterhof & Todorov, 2008). All of these 120 pictures depicted bald, Caucasian males whose facial appearance differed in the level of trustworthiness. Specifically, we used 45 faces with trustworthy facial expressions and 45 faces with untrustworthy facial expressions to induce expectations of trustworthiness and untrustworthiness, respectively. In addition, we selected 30 faces with neutral facial expressions for practice trials. All faces were

generated using FaceGen Modeller Version 3.1 and 3.2 (Singular Inversions, 2005, 2007) according to the trustworthiness computer model developed by Oosterhof and Todorov (2008). Examples for each of the three face categories can be found in Figure 1.

Furthermore, a total of 198 German nouns and adjectives that we thought suitable to describe person characteristics were selected from two freely accessible databases: the Berlin Affective Word List Reloaded (BAWL-R; Vö et al., 2009) and the Age-Dependent Evaluations of German Adjectives database (AGE; Grühn & Smith, 2008a, 2008b). In a second step, 136 undergraduate students ($M_{age}=23.9$, $SD=9.72$, 77% female) rated the trustworthiness of each of the 198 words on a scale from 1 [very strongly associated with untrustworthiness] to 7 [very strongly associated with trustworthiness]. Based on the ratings of this norming study, we selected 20 trustworthy words with a mean trustworthiness of 5.66 ($SD=0.33$), 20 untrustworthy words with a mean trustworthiness of 1.71 ($SD=0.30$), and 80 neutral words ($M=4.08$, $SD=0.51$) as final stimulus material. Words associated with trustworthiness were for example “honest” or “just”, words associated with untrustworthiness were for example “greedy” and “selfish”, and neutral words were for example “smoker” and “popular”. In a final step, these 120 words were sorted into 30 combinations consisting of four words each. Combinations always entailed one trustworthy and three neutral words (trustworthy word condition; $n=10$), one untrustworthy and three neutral words (untrustworthy word condition; $n=10$) or one trustworthy, one untrustworthy, and two neutral words (competition condition; $n=10$). Importantly, words within one combination were matched with regard to word length (i.e., number of letters) and word frequency (i.e., frequency of appearance per million words); these features have been shown to influence word processing (e.g., Vö et al., 2009).² In addition, each individual word appeared only in one combination and all words were capitalized for standardization purposes.

² We also planned to match the words with regard to word valence and arousal. However, this turned out to be difficult because trustworthiness, valence, and arousal ratings were highly correlated ($r = .57-.82$). Not

Design

We used a within-subjects design with face condition (trustworthy vs. untrustworthy face) and word condition (trustworthy word vs. untrustworthy word vs. competition condition) as factors, resulting in six experimental conditions (see Figure 2). Each trial consisted of the presentation of a face together with one randomly chosen word combination. Therefore, trials constituted either a match between word condition and face condition (e.g., untrustworthy word and untrustworthy face), a mismatch between word condition and face condition (e.g., untrustworthy word and trustworthy face), or both (e.g., competition condition and trustworthy face).³

In total, the experiment consisted of 300 trials that were organized in 10 blocks with 30 trials each. Importantly, only trustworthy *or* untrustworthy faces were used in each block, resulting in 50% untrustworthy and 50% trustworthy face blocks. The order in which these blocks were presented was counterbalanced across participants. In addition, each block contained all 30 word combinations and 30 different faces in a random order, but combinations and faces were used repeatedly across blocks.

Procedure

The design and methods of this study were approved by a local ethics committee. Participants were tested individually in a session that took about 1.5 hours and gave informed consent before the experiment began. On arrival, visual acuity was assessed using an Oculus Binoptometer 3 (a visual acuity of 0.8 was required for participation), and impairments in stereo vision and color vision were ruled out. Participants were then informed that the study

entirely surprising, we found untrustworthy words to be more negative and more arousing than words with neutral trustworthiness ratings, while trustworthy words were more positive and less arousing. As a consequence, words within one combination differed not only in trustworthiness levels but also in valence and arousal.

³ However, results in the two competition conditions were inconsistent and will thus not be referred to further (see Appendix A for an overview of the results). The lack of findings in these conditions may be due to the fact that conflicting information is given (a face is always presented with a trustworthy and an untrustworthy word simultaneously, which might have cancelled each other out).

consisted of a computer-based eye tracking experiment that was ostensibly designed to investigate “the relation between eye movements and memory in the context of different feature configurations”. We used this cover story to disguise the actual purpose of the study and also implemented a memory task to make it more plausible. For this reason, the students were told that they would see faces together with words, and that after every fifth trial, a recognition test would take place. In this recognition test, participants had to categorize a word as either “old” (i.e., contained within the last five trials) or “new”, and feedback on task performance was given at regular intervals.⁴

At the beginning of the experiment, participants were seated in front of the computer screen. After calibration of the eye tracker, participants were first familiarized with the task in a practice block consisting of 30 trials⁵, followed by 10 experimental blocks. Between blocks, participants were informed about their recognition accuracy in the “memory task” and had the opportunity to take short breaks before the experiment was continued. Within experimental blocks, each trial started with a fixation dot surrounded by a thin line presented at screen center. After participants successfully fixated this dot, the trial started and the thin line disappeared. After 500 milliseconds (ms), a picture of a male face with a trustworthy or untrustworthy facial expression replaced the fixation dot and was presented for 2000 ms, before four words complemented the face for another 3000 ms. Each word was presented in one quadrant of the display, so that every word had the same distance to the face at screen center. The assignment of the four words of each combination to the four quadrants was randomized for every trial. In addition, the font size of the words was adjusted (Arial 36, bold) so that the words were peripherally recognizable—at least to some extent—when the

⁴ We included this memory task not only to distract participants from the actual purpose of the study but also to make sure that the presented words were read and processed in a comparable manner. However, participants’ recognition accuracy was of no further interest.

⁵ Practice trials were identical to the subsequent experimental trials but only contained faces with neutral facial expressions and words with neutral trustworthiness ratings not used in the main experiment.

face at screen center was fixated. After the presentation of the face and the words, a blank screen appeared for 500 ms, before the next trial started (see Figure 3 for a schematic overview of the trial procedure).

After the main experiment, participants took part in two additional tasks that we included for exploratory purposes: first, 30 additional trials were implemented that also consisted of the presentation of a face and four words like in the main experiment.⁶ This time, however, participants were instructed to rate the likeability of each face on a scale from 1 [very dislikeable] to 7 [very likeable]. Furthermore, a nonverbal working memory test (PEBL Corsi block-tapping task; Mueller & Piper, 2014) was conducted.

Finally, participants were probed for suspicion and answered a short follow-up survey about the experimental tasks and the strategies they applied during the tasks. Next, victim sensitivity was assessed with ten items from the Justice Sensitivity Inventory (Cronbach's $\alpha=.84$; Schmitt et al., 2010), that were rated on a six-point Likert scale from 0 [not at all true] to 5 [absolutely true]. Example items are "I ruminate for a long time when other people are treated better than me", and "It makes me angry when others receive a reward that I have earned". In addition, we assessed four personality scales as possible covariates, namely general trust (with a German version of the General Trust Scale by Yamagishi & Yamagishi, 1994), neuroticism (BFI-10; Rammstedt & John, 2007) and the other three facets of justice sensitivity (observer, beneficiary, and perpetrator sensitivity; Schmitt et al., 2010). Finally, participants were thanked for their participation and fully debriefed.

Dependent Variables

We used the software EyeLink Data Viewer (SR Research Ltd.) to pre-process the eye tracking data and IBM SPSS Statistics Version 25 for subsequent statistical analyses. The main outcome variables in this study were different eye tracking parameters that we

⁶ For this additional block, we used the same word combinations as before but selected fifteen new trustworthy and fifteen new untrustworthy faces from our pool of male faces.

calculated separately for each of the four words in each trial. Neutral words, however, were not part of our hypotheses and will thus not be referred to any further.

Attention allocation was operationalized via several eye tracking parameters (see Holmqvist et al., 2011; Süßenbach et al., 2012). More specifically, we measured (1) how long participants fixated the word when first landing on it (“first fixation duration”), (2) how long the word was fixated in total (“dwell time”), and (3) how often participants fixated the respective word per trial (“fixation count”). In addition, we also measured (4) which word was fixated first in each trial (“destination of the first saccade”). These dependent variables represent slightly different “stages” of attention and visual information processing and reflect earlier/faster (destination of the first saccade, first fixation duration) as well as later/slower (dwell time, fixation count) cognitive processes (see Holmqvist et al., 2011). In reading research, for example, the duration of the first fixation is often defined as a measure of identification or (lexical) processing time, while dwell time—i.e., the total amount of time spent fixating on a word—is sought to reflect higher integrational and evaluative processes as well (Holmqvist et al., 2011; Inhoff & Radach, 1998; Rayner, 1998). Therefore, effects of violations and confirmations of (un)trustworthiness expectations may differ between these four dependent variables (Inhoff & Radach, 1998).

Results

We excluded trials from analyses in which no word was fixated, which applied to 32 trials (0.001%) in total. Mean values on each of the four dependent variables described above are shown in Table 1, broken down by experimental condition. To compare attention allocation between words, we calculated “attentional bias scores” for each dependent variable for trustworthy and untrustworthy faces. These bias scores—or difference scores—were calculated across trustworthy and untrustworthy word conditions. More specifically, mean scores for face-congruent words were subtracted from mean scores for face-incongruent

words. Thus, positive scores indicate preferential attention towards face-incongruent words relative to face-congruent words, while negative scores reflect an attentional bias away from face-incongruent words. Because bias scores were calculated separately for each of the four dependent variables described above (first fixation duration, dwell time, fixation count, destination of the first saccade), broken down by face types (untrustworthy vs. trustworthy faces), this procedure resulted in 8 attentional bias scores in total.

To test our hypotheses, bivariate correlations between attentional bias scores and victim sensitivity were probed for significance (see below). Although we will focus mainly on these correlations in the following results sections, we also used multilevel modeling (i.e., mixed models) to analyze the data with regard to first fixation duration, dwell time, and fixation count. Here, random intercepts were modeled to account for the nested data structure. More specifically, two fixed level-1 predictors were included for face type (untrustworthy = -0.5, trustworthy = +0.5) and word type (untrustworthy = -0.5, trustworthy = +0.5), respectively. At level 2, victim sensitivity (z-standardized) as well as the respective interaction terms were entered. Three-way interaction effects between face type, word type, and victim sensitivity were probed for significance. Results of these analyses are reported in the subsequent footnotes. Detailed results are reported in Appendix B.

Attentional Biases in the Trustworthy and Untrustworthy Word Conditions

Overall, attentional biases were only observed in some, but not in all dependent variables (see Table 1). For untrustworthy faces, we found significant positive difference scores for fixation count ($t(68)=5.275, p < .001, d=0.63, 95\% \text{ CI } [0.37, 0.89]$) and dwell time ($t(68)=4.668, p < .001, d=0.56, 95\% \text{ CI } [0.31, 0.81]$), illustrating that after the presentation of untrustworthy faces, face-incongruent trustworthy words were fixated more often and for a longer time than face-congruent untrustworthy words. In the trustworthy face condition, however, a negative bias score was visible with regard to fixation count ($t(68)= -4.183, p <$

.001, $d = -0.51$, 95% CI [-0.76, -0.25]): trustworthy words—although confirming the facial expression—were again fixated more often than untrustworthy words. Thus, these “main effects” speak for an attentional bias towards positive, trustworthiness-related words in general.⁷

Hypothesis Tests: Correlations with Victim Sensitivity

To test the assumption that victim sensitivity predicts attention allocation towards expectancy violations, we correlated attentional bias scores with victim sensitivity in a second step. These correlations are reported in Table 2. In addition, a graphical representation of predicted means (from the multilevel analyses) in first fixation duration, dwell time, and fixation count—as a function of victim sensitivity, face type, and word type—is shown in Figure 4.

First Fixation Duration

In line with our hypothesis, we found a positive correlation between victim sensitivity and the attentional bias score for first fixation durations in the context of untrustworthy faces ($r(67) = .23$, $p = .06$, two-tailed, 95% CI [-0.01, 0.44]⁸). Thus, victim sensitivity predicted longer initial fixations on face-incongruent trustworthy words than on face-congruent untrustworthy words after untrustworthy faces had been presented. In contrast, we did not find a significant correlation in the context of trustworthy faces ($r(67) = .02$, $p = .85$, two-tailed, 95% CI [-0.22, 0.26]), implying that victim sensitivity did not predict longer first fixations on face-incongruent words in the context of trustworthy faces. In other words, no attentional bias was found for negative expectancy violations.⁹

⁷ In the multilevel analyses, we found a significant main effect of word type ($p_s < .001$), while the main effect of face type and the face type x word type interactions did not reach significance ($p_s > .10$). These results thus confirm that—across face conditions—trustworthy words were fixated more often and for a longer time than untrustworthy words.

⁸ Note: in the context of correlation analyses, p -values $< .10$ are interpreted as significant when hypotheses are directional and one-tailed testing is therefore warranted.

⁹ In our multilevel analysis, the three-way interaction between victim sensitivity, face type, and word type did not reach significance ($p = .15$). However, in separate analyses for the two face conditions, the expected victim

Dwell Time

In contrast to first fixation duration, victim sensitivity was uncorrelated ($p = .47$) with dwell time differences in the context of untrustworthy faces. Therefore, victim sensitivity did not predict longer dwell times on face-incongruent (vs. face-congruent) words in this facial context. However, we did observe a significant positive correlation between victim sensitivity and dwell time differences in the context of trustworthy faces ($r(67) = .27, p = .03$, two-tailed, 95% CI [0.04, 0.48]), suggesting that participants high (vs. low) in victim sensitivity spent more time fixating face-incongruent untrustworthy words compared to face-congruent trustworthy words.¹⁰ However, as Figure 4 shows, this effect was mainly driven by a strong negative correlation between victim sensitivity and dwell times on trustworthy words. Put differently: while people low in victim sensitivity allocated preferential attention toward face-congruent trustworthy words in the context of trustworthy faces, this consistency or positivity effect disappeared (and was even reversed) with increasing victim sensitivity.

Fixation Count

Victim sensitivity was correlated with the attentional bias score for fixation count both in the context of untrustworthy faces ($r(67) = -.25, p = .04$, two-tailed, 95% CI [-0.46, -0.01]) as well as in the context of trustworthy faces ($r(67) = .26, p = .03$, two-tailed, 95% CI [0.03, 0.47]). Contrary to our hypothesis, the negative correlation for untrustworthy faces implied *fewer* fixations on (face-incongruent) trustworthy words for people high (vs. low) in victim sensitivity, that is, an attentional bias away from positive expectancy violations. However, the positive correlation in the context of trustworthy faces shows that higher victim sensitivity

sensitivity x word type interaction was marginally significant for untrustworthy faces ($p = .075$), while no such interaction was found for trustworthy faces ($p = .83$).

¹⁰ The results of the multilevel analysis confirm these findings. In accordance with the correlational analyses, we found a significant victim sensitivity x face type x word type three-way interaction effect ($p = .04$). Following up on this significant interaction, separate analyses for each face condition demonstrated a significant interaction between VS and word type only for trustworthy faces ($p = .02$), but not for untrustworthy faces ($p = .50$).

was also associated with fewer fixations on trustworthy words after trustworthy faces had been presented.¹¹ In other words, trustworthy words were always fixated more often than untrustworthy words when victim sensitivity was low, but this “positivity bias” (which corresponds to the main effect of word type found for fixation count) disappeared with increasing victim sensitivity. Thus, whereas *victim-insensitive* participants were more inclined to fixate on trustworthy words than on untrustworthy words (irrespective of whether these words were presented after a trustworthy or an untrustworthy face), *victim-sensitive* individuals tended to fixate equally often on both word types in both facial contexts. Therefore, these effects seem to reflect attention allocation towards (or away from) specific word content rather than reactions to expectancy violation vs. confirmation.

Destination of the first Saccade

Bias scores in this dependent variable were uncorrelated with victim sensitivity in both face conditions (all $p_s > .23$).

In sum, our findings show that victim sensitivity (VS) is associated with preferential attention towards positive expectancy violations. Participants high (vs. low) in VS showed longer initial fixations on trustworthy words than on untrustworthy words after untrustworthy faces had been presented. However, participants high in VS also showed a general tendency to fixate trustworthy words less often, independent of the facial context. This pattern of results was somewhat unexpected but (1) it seemed to reflect an attentional bias to specific word content rather than to expectancy violations (because the effect appeared in both face conditions), and (2) it may be explained by the fact that trial durations were held constant in our study. More specifically, we hypothesized that because word presentation was limited to 3000 ms, the number of fixations might be negatively related to first fixation durations (c.f.

¹¹ Accordingly, the results of the multilevel analysis showed a significant victim sensitivity x word type interaction across face types ($p = .002$), while the VS x word type x face type three-way interaction was not significant ($p = .95$).

Holmqvist et al., 2011). Indeed, exploratory correlation analyses revealed a negative relation between the bias scores in first fixation duration and fixation count ($r(67) = -.36, p = .002$, two-tailed, 95% CI $[-0.55, -0.14]$). This finding implies that when the initial fixation on the face-incongruent word (relative to the face-congruent word) was longer, fewer fixations were credited to this word (relative to the face-congruent word) in total—presumably because a longer initial fixation might have allowed for a more profound processing already.

In contrast to positive expectancy violations, our results indicate that VS was not associated with an attentional bias towards negative expectancy violations. Although we did find significant correlations of VS with dwell time and fixation count differences in the context of trustworthy faces, these seemed to be mainly driven by a positivity bias associated with low VS, which disappeared with increasing VS (cf. Figure 4). Thus, whereas victim-insensitive participants showed longer dwell times and more fixations on trustworthy words than on untrustworthy words in this facial context, victim-sensitive individuals fixated about equally long (and equally often) on both word types.

Discussion

In the present research, we examined how victim-sensitive individuals process social information related to trustworthiness or untrustworthiness, and, more specifically, how they process positive and negative expectancy violations in this context. In line with our primary hypothesis, victim sensitivity (VS) predicted attention allocation towards positive expectancy violations: after the presentation of untrustworthy faces, victim-sensitive individuals (but not victim-insensitive individuals) tended to fixate face-incongruent trustworthy words longer than face-congruent untrustworthy words when first landing on it, which suggests a deeper processing of cues that positively violate negative initial expectations (Holmqvist et al., 2011). This result corroborates our assumption that violations of untrustworthiness expectations are especially motivationally relevant for victim-sensitive persons (because of

their strong need to trust, c.f. Gollwitzer et al., 2013), and therefore receive preferential attention. In addition, this result is in line with previous findings showing that people high (vs. low) in VS are more likely to (a) remember information suggesting that an initially untrustworthy target person may not be so untrustworthy after all, and (b) update their expectations more readily in such cases of positive expectation violations (Süssenbach et al., 2016). Possibly, these findings can be explained by the effects observed in the present study: victim-sensitive people preferentially attend to stimuli violating their negative expectations; and therefore, these stimuli are stored in memory (Chun et al., 2011; Chun & Turk-Browne, 2007).

Notably, VS not only predicted longer first fixations on positive expectancy violations, but also *fewer* fixations on these words. At first glance, the reduced number of fixations seems to contradict the notion that, for victim-sensitive individuals, violations of untrustworthiness expectations are rather attended than confirmations of such expectations. However, high VS was associated with fewer fixations on trustworthy words in both facial contexts (i.e., while victim-insensitive individuals fixated more on trustworthy than on untrustworthy words in both face conditions, victim-sensitive participants did show no such “positivity bias”). Therefore, differences in this dependent variable seem to reflect attention allocation towards (or away from) specific word content, independent of any expectations the participants might have held.

Moreover, characteristics of the “memory task” that we used in our study may have contributed to the lower fixation count for positive expectancy violations. As mentioned above, we included this task for two reasons: (1) to increase the credibility of the cover story and (2) to ensure that participants would read and semantically process the words presented in each trial. As part of this task, a recognition test was implemented after every fifth trial, in which participants were asked to identify a presented word as “old” (i.e., contained in one of

the last five trials) or “new”. Thus, participants were motivated to fixate each of the four words in each trial at least once to successfully memorize them. In addition, the task was associated with time pressure, as the words were presented for only 3000 ms in total. Therefore, the memory task might have contributed to the lower number of fixations on expectancy-violating trustworthy words: probably, victim-sensitive persons showed fewer fixations on these words (relative to expectancy-confirming untrustworthy words), because they were sufficiently processed during the prolonged first fixation and because the memory task demanded to attend the other words as well. Previous research confirms an inverse relationship between number of fixations and fixation durations when trial durations are held constant (Holmqvist et al., 2011), and the difference scores of fixation count and first fixation duration were negatively correlated in our study as well.

Importantly, although the memory task might have affected our results in some ways, it should be noted that its implementation resulted in a more *conservative* testing of our hypothesis. Because of this task, participants were not motivated to focus their attention on one word but to try to attend to *all* words. As a consequence, differences in attention allocation between face-congruent and face-incongruent words should be small, and therefore harder to detect. As both nature and difficulty of the task at hand have been shown to influence cognitive processes and eye movements (Rayner 1998, 2009), other experimental tasks and instructions—such as free reading or search tasks—would probably result in different viewing patterns. Future studies should hence compare attention allocation under different instructions to see whether our findings generalize to other experimental tasks and settings.

Characteristics of the materials and instructions used in our study might also be responsible for the lack of effects on the “destination of the first saccade”, which can be seen as an indicator of early attention capture. Contrary to our hypothesis, results showed that

expectancy-violating words were not fixated earlier in the trial than expectancy-confirming words, irrespective of face or word condition or VS. One plausible reason for this null result might be the tendency of most participants to scan all the words presented on the screen in the order in which they are usually processed (i.e., from the upper left to the lower right side in Latin script). These habits might have overshadowed any “destination of the first saccade” effects. Furthermore, it is also possible that expectancy-violating words did not grab attention because participants were not able to adequately process the words in parafoveal vision when fixating on the face at screen center (although we made sure to use an appropriate font size). Future research should therefore consider using pictorial material to avoid these difficulties.

Negative Expectancy Violations

In accordance with the findings reported by Süssenbach et al. (2016), VS was not associated with preferential attention towards *negative expectancy violations* in our study. Notably, we did find significant correlations between VS and the difference scores in dwell time and fixation count in the context of trustworthy faces. However, effects in fixation count were—as already described—not limited to one face type and therefore presumably represent effects of specific word content rather than effects of expectancy violations. In addition, visualizations of the relationships implied that the correlations were mainly driven by victim-insensitive individuals’ preferential attention for trustworthy words (i.e., words confirming the positive expectation). Put differently: while victim-insensitive individuals showed both longer dwell times and a larger number of fixations on expectation-confirming trustworthy words than on expectation-violating untrustworthy words, victim-sensitive people attended both word types about equally long and often. Thus, low VS was related to a positivity bias that disappeared with increasing victim sensitivity, but high VS was not associated with an attentional bias toward negative expectancy violations.

Limitations

In the present study, we used a newly developed paradigm to provide first evidence for the assumption that victim sensitivity predicts attention allocation towards information that violates untrustworthiness expectations. However, in addition to the strengths of the study there are also some limitations to consider.

First, the words used in the present experiment—although matched on important dimensions like word length and word frequency—differed not only with regard to trustworthiness, but also with regard to valence and arousal. In fact, ratings on these three dimensions showed high intercorrelations, and while untrustworthy words were found to be negative and highly arousing, trustworthy words tended to be positive and less arousing. In addition, the same applied to the facial stimuli that we selected. As can be seen in Figure 1, the faces that we used to induce trustworthiness expectations resemble friendly-looking, smiling faces that signal happiness. The untrustworthy faces, on the other hand, have narrowed eyes and look much more “grumpy”, which results in a more aggressive or hostile appearance. Therefore, trustworthiness judgements of faces correlate highly with general valence evaluations as well (see Oosterhof & Todorov, 2008; Todorov et al., 2008). Thus, it seems to be almost impossible to disentangle these features entirely.

Second, we manipulated expectations and the motivational state of our participants in a very indirect way. In our study, pictures of untrustworthy looking male targets were used to threaten the need to trust and to induce expectations of untrustworthiness, but these confrontations never resulted in actual interactions or consequences. Thus, participants were mere observers in our paradigm who did not experience victimization or exploitation per se. For this reason, future studies might want to threaten victim-sensitive individuals’ need to trust in a more salient way, to be able to draw conclusive inferences about the motivational concerns of people high in victim sensitivity. One possibility to achieve this could be to thwart the need to trust through a direct experience of exploitation, for example in the context

of a trust game (e.g., Berg et al., 1995; Gollwitzer & Rothmund, 2011). More specifically, one could think of an experimental design in which one half of the participants is being exploited by an interaction partner, whereas the other half experiences a neutral or positive social exchange. In our opinion, a manipulation like this should lead to even stronger effects than the ones obtained in the present study, and it would also provide additional insights into the underlying processes of attention allocation. Therefore, we consider it a useful avenue for future research.

Conclusion

The current findings provide first evidence that victim sensitivity influences the visual processing of social information related to (un)trustworthiness. Our results demonstrate that victim sensitivity predicts attention allocation towards positive expectancy violations, which are considered to be especially motivationally relevant for victim-sensitive individuals. Importantly, this attentional bias was not visible in all eye tracking measures: (high) victim sensitivity was only associated with longer first fixation durations on words that violated (vs. confirmed) untrustworthiness expectations, but not with longer dwell times or a higher fixation count. It therefore seems that victim sensitivity has an impact on earlier stages of the attentional processing of positive expectancy violations (prolonging the first fixation duration), while later stages (fixation count, dwell times) are rather unaffected. Furthermore, our study suggests that victim sensitivity is not associated with an attentional bias towards negative expectancy violations. In conclusion, the present research sheds light on the cognitive processes by which individuals high vs. low in victim sensitivity process violations and confirmations of (un)trustworthiness expectancies, and it adds to our understanding of the motivational concerns underlying victim sensitivity.

Declarations

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Ethics Committee of the department of psychology of Philipps-Universität Marburg.

References

- Aly, M., & Turk-Browne, N. B. (2016). Attention promotes episodic encoding by stabilizing hippocampal representations. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), E420–E429.
<https://doi.org/10.1073/pnas.1518931113>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. <https://doi.org/10.1006/game.1995.1027>
- Bergmann, N., Koch, D., & Schubö, A. (2019). Reward expectation facilitates context learning and attentional guidance in visual search. *Journal of Vision*, *19*(3), 1–18.
<https://doi.org/10.1167/19.3.10>
- Brosch, T., & Van Bavel, J. J. (2012). The flexibility of emotional attention: Accessible social identities guide rapid attentional orienting. *Cognition*, *125*(2), 309–316.
<https://doi.org/10.1016/j.cognition.2012.07.007>
- Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*, 613–625.
<https://doi.org/10.1038/nrn2459>
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.
<https://doi.org/10.1146/annurev.psych.093008.100427>
- Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184.
<https://doi.org/10.1016/j.conb.2007.03.005>
- Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II: Case study: A computational theory of social exchange. *Ethology and Sociobiology*, *10*(1–3), 51–97. [https://doi.org/10.1016/0162-3095\(89\)90013-7](https://doi.org/10.1016/0162-3095(89)90013-7)

- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. [https://doi.org/10.1016/0042-6989\(95\)00294-4](https://doi.org/10.1016/0042-6989(95)00294-4)
- DeWall, C. N., Maner, J. K., & Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: Selective attention to signs of acceptance. *Journal of Personality and Social Psychology*, *96*(4), 729–741. <https://doi.org/10.1037/a0014634>
- Dietze, P., & Knowles, E. D. (2016). Social class and the motivational relevance of other human beings: Evidence from visual attention. *Psychological Science*, *27*(11), 1–11. <https://doi.org/10.1177/0956797616667721>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Feldmann-Wüstefeld, T., Brandhofer, R., & Schubö, A. (2016). Rewarded visual items capture attention only in heterogeneous contexts. *Psychophysiology*, *53*(7), 1063–1073. <https://doi.org/10.1111/psyp.12641>
- Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences*, *36*(5), 1015–1029. [https://doi.org/10.1016/S0191-8869\(03\)00197-1](https://doi.org/10.1016/S0191-8869(03)00197-1)
- Gerlach, T. M., Allemand, M., Agroskin, D., & Denissen, J. J. A. (2012). Justice sensitivity and forgiveness in close interpersonal relationships: The mediating role of mistrustful, legitimizing, and pro-relationship cognitions. *Journal of Personality*, *80*(5), 1373–1413. <https://doi.org/10.1111/j.1467-6494.2012.00762.x>
- Gollwitzer, M., & Rothmund, T. (2009). When the need to trust results in unethical behavior:

- The Sensitivity to Mean Intentions (SeMI) model. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 135–152). Information Age Publishing.
- Gollwitzer, M., & Rothmund, T. (2011). What exactly are victim-sensitive persons sensitive to? *Journal of Research in Personality, 45*(5), 448–455.
<https://doi.org/10.1016/j.jrp.2011.05.003>
- Gollwitzer, M., Rothmund, T., Alt, B., & Jekel, M. (2012). Victim sensitivity and the accuracy of social judgments. *Personality and Social Psychology Bulletin, 38*(8), 975–984. <https://doi.org/10.1177/0146167212440887>
- Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality, 43*(6), 999–1005. <https://doi.org/10.1016/j.jrp.2009.07.003>
- Gollwitzer, M., Rothmund, T., & Süßenbach, P. (2013). The Sensitivity to Mean Intentions (SeMI) model: Basic assumptions, recent findings, and potential avenues for future research. *Social and Personality Psychology Compass, 7*(7), 415–426.
<https://doi.org/10.1111/spc3.12041>
- Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical effects of justice sensitivity perspectives on prosocial and antisocial behavior. *Social Justice Research, 18*(2), 183–201. <https://doi.org/10.1007/s11211-005-7368-1>
- Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences, 19*(9), 524–533.
<https://doi.org/10.1016/j.tics.2015.07.009>
- Grühn, D., & Smith, J. (2008a). Characteristics for 200 words rated by young and older adults: Age-dependent evaluations of German adjectives (AGE). *Behavior Research Methods, 40*(4), 1088–1097. <https://doi.org/10.3758/BRM.40.4.1088>

- Grühn, D., & Smith, J. (2008b). Gruhn-Smith_2008_AGE.txt. Retrieved November 16, 2017 from Psychonomic Society Web Archive: <http://www.psychonomic.org/ARCHIVE/>
- Heuer, A., & Schubö, A. (2018). Separate and combined effects of action relevance and motivational value on visual working memory. *Journal of Vision, 18*(5), 1–14. <https://doi.org/10.1167/18.5.14>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Elsevier.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons & M. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97–135). Erlbaum.
- Lavie, N., & Dalton, P. (2014). Load theory of attention and cognitive control. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 56–75). Oxford University Press.
- Maltese, S., Baumert, A., Schmitt, M. J., & MacLeod, C. (2016). How victim sensitivity leads to uncooperative behavior via expectancies of injustice. *Frontiers in Psychology, 6*, 2059. <https://doi.org/10.3389/fpsyg.2015.02059>
- Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology, 17*(2), 119–128. [https://doi.org/10.1016/0162-3095\(95\)00131-X](https://doi.org/10.1016/0162-3095(95)00131-X)
- Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL)

- and PEBL Test Battery. *Journal of Neuroscience Methods*, 222, 250–259.
<https://doi.org/10.1016/j.jneumeth.2013.10.024>
- Oda, R. (1997). Biased face recognition in the Prisoner's Dilemma Game. *Evolution and Human Behavior*, 18(5), 309–315. [https://doi.org/10.1016/S1090-5138\(97\)00014-7](https://doi.org/10.1016/S1090-5138(97)00014-7)
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA*, 105(32), 11087–11092.
<https://doi.org/10.1073/pnas.0805664105>
- Proulx, T., Slegers, W., & Tritt, S. M. (2017). The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology*, 70, 69–79. <https://doi.org/10.1016/j.jesp.2016.12.003>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506,
<https://doi.org/10.1080/17470210902816461>
- Rief, W., Glombiewski, J. A., Gollwitzer, M., Schubö, A., Schwarting, R., & Thorwart, A. (2015). Expectancies as core features of mental disorders. *Current Opinion in Psychiatry*, 28(5), 378–38. <https://doi.org/10.1097/YCO.0000000000000184>
- Rusting, C. L. (1998). Personality, mood, and cognitive processing of emotional information: Three conceptual frameworks. *Psychological Bulletin*, 124(2), 165–196.
<https://doi.org/10.1037/0033-2909.124.2.165>

- Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*, 211–238.
<https://doi.org/10.1007/s11211-010-0115-2>
- Singular Inversions. (2005). *FaceGen main software development kit* (version 3.1).
Singular Inversions.
- Singular Inversions. (2007). *FaceGen main software development kit* (version 3.2).
Singular Inversions.
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Süssenbach, P., Bohner, G., & Eyssel, F. (2012). Schematic influences of rape myth acceptance on visual information processing: An eye-tracking approach. *Journal of Experimental Social Psychology, 48*(3), 660–668.
<https://doi.org/10.1016/j.jesp.2011.12.005>
- Süssenbach, P., Gollwitzer, M., Mieth, L., Buchner, A., & Bell, R. (2016). Trustworthy tricksters: Violating a negative social expectation affects source memory and person perception when fear of exploitation is high. *Frontiers in Psychology, 7*, 1–12.
<https://doi.org/10.3389/fpsyg.2016.02037>
- Theeuwes, J. (2018). Visual selection: Usually fast and automatic; seldom slow and volitional. *Journal of Cognition, 1*(1), 1–15. <https://doi.org/10.5334/joc.13>
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833.
<https://doi.org/10.1521/soco.2009.27.6.813>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation

- of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455–460.
<https://doi.org/10.1016/j.tics.2008.10.001>
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538. <https://doi.org/10.3758/BRM.41.2.534>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592–598.
<https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research*, 47(7), 955–964.
<https://doi.org/10.1016/j.visres.2006.11.025>
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, 18(2), 129–166. <https://doi.org/10.1007/BF02249397>



Fig.1 Examples of untrustworthy (left), neutral (middle), and trustworthy (right) faces.

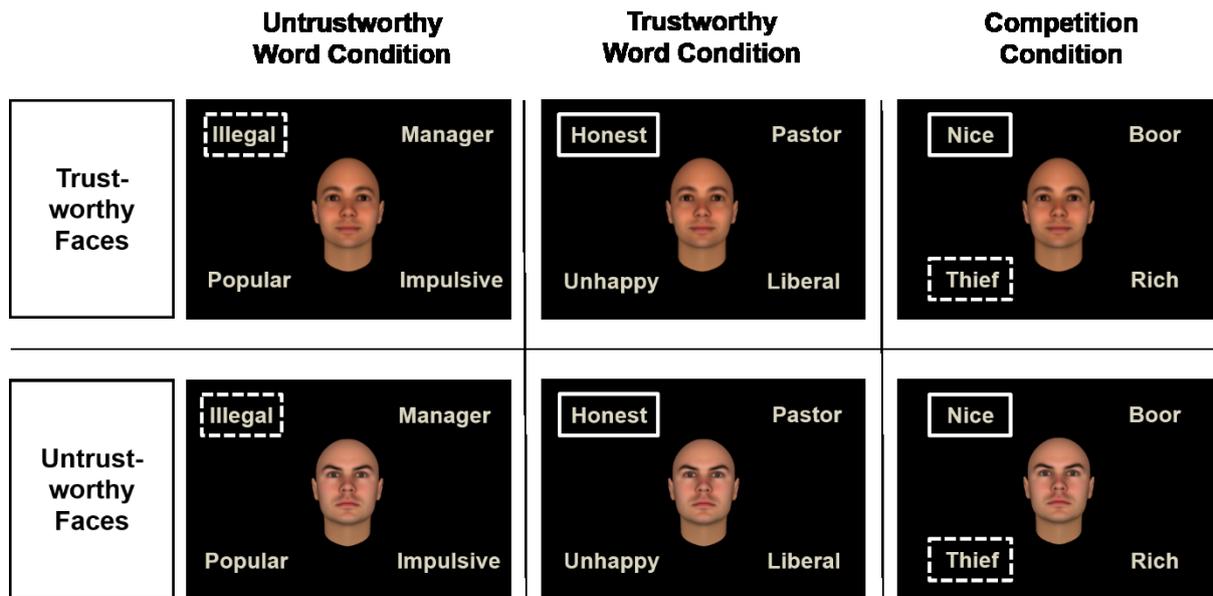


Fig.2 Experimental conditions. Either a trustworthy (first line) or an untrustworthy face (second line) was presented at screen center. The face was complemented by either one untrustworthy and three neutral words (left column), one trustworthy and three neutral words (middle column), or one trustworthy, one untrustworthy, and two neutral words (right column). In this figure, trustworthy and untrustworthy words are highlighted by solid and dashed boxes, respectively, but these markings were not shown in the experiment.

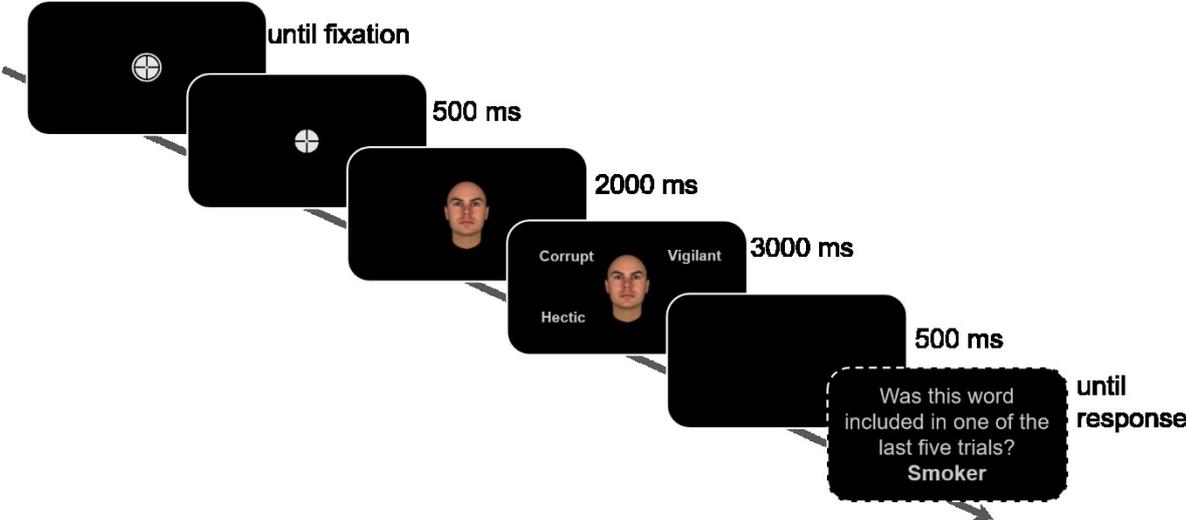


Fig.3 Schematic overview of one experimental trial.

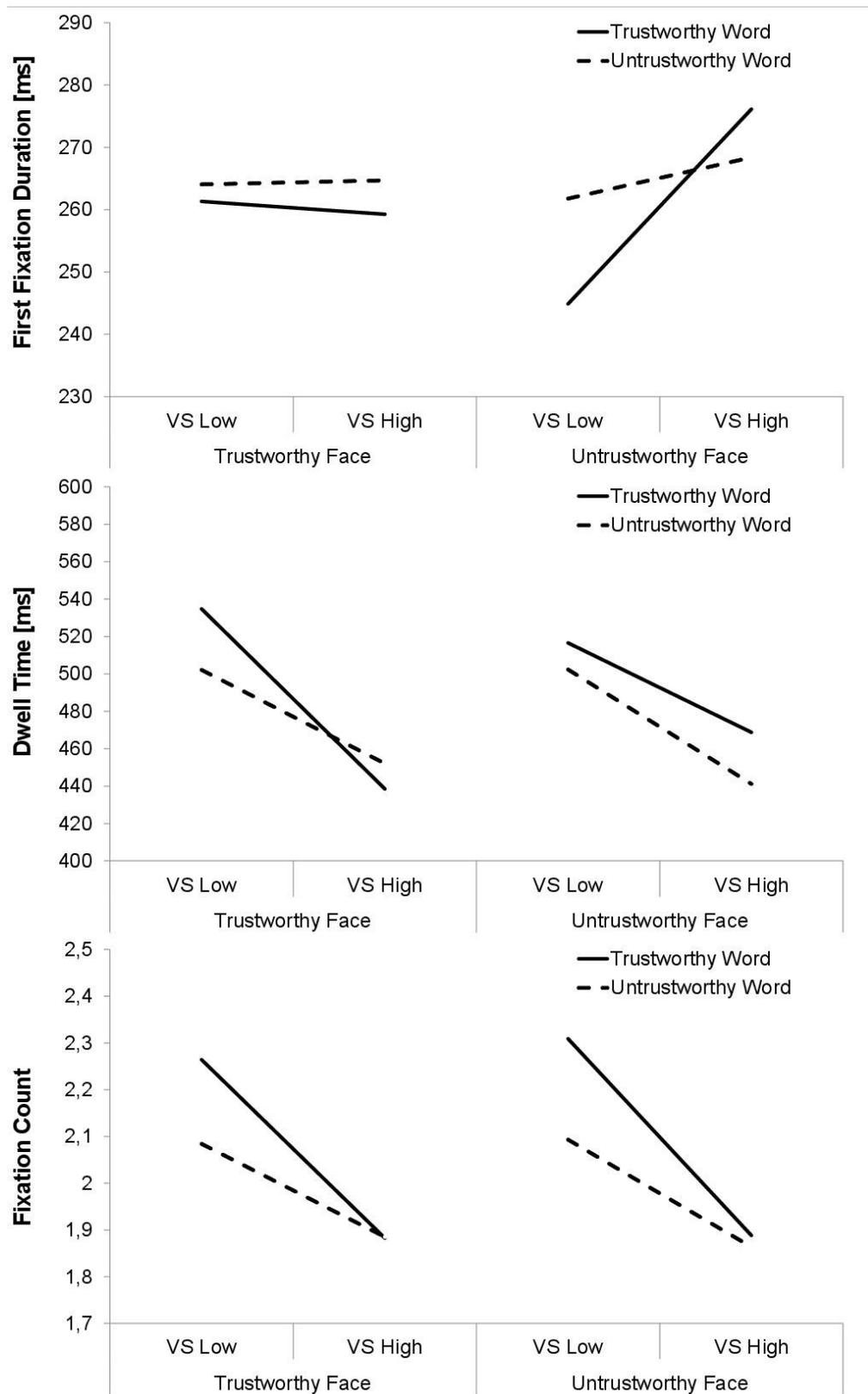


Fig.4 Predicted means in first fixation duration, dwell time, and fixation count (obtained from the multilevel analyses) visualizing the face type × word type interaction effects for participants low vs. high in victim sensitivity (± 2 SD from the sample mean).

Table 1*Means and Standard Deviations of Eye Movement Variables for Relevant Words**Trustworthy and Untrustworthy Word Conditions*

Word	First fixation duration [ms]	Dwell time [ms]	Fixation count	Destination of the first saccade [%]
U face				
U word	265 (45)	472 ^a (70)	1.98 ^a (0.33)	24.75 (6.25)
T word	261 (47)	493 ^a (79)	2.10 ^a (0.41)	24.61 (6.15)
T face				
U word	264 (45)	477 (71)	1.99 ^a (0.36)	24.35 (6.09)
T word	260 (48)	487 (78)	2.07 ^a (0.39)	25.48 (6.12)

Note. U face = untrustworthy face, T face = trustworthy face, U word = untrustworthy word, T word = trustworthy word. Values reported for “destination of the first saccade” represent the percentage of trials in which the word was fixated first. Standard deviations are shown in parentheses.

^a Significant difference between trustworthy and untrustworthy words in one face condition ($p < .001$, two-tailed).

Table 2*Correlations of Attentional Bias Scores with Victim Sensitivity**Trustworthy and Untrustworthy Word Conditions*

Bias score	VS	<i>p</i>	95% CI	
			<i>LL</i>	<i>UL</i>
U face				
First fixation duration	.23 ⁺	.055	-.007	.443
Dwell time	.09	.466	-.150	.320
Fixation count	-.25*	.037	-.459	-.014
Destination of the first saccade	-.06	.600	-.293	.179
T face				
First fixation duration	.02	.846	-.218	.255
Dwell time	.27*	.025	.036	.476
Fixation count	.26*	.032	.025	.468
Destination of the first saccade	.15	.228	-.090	.373

Note. U face = untrustworthy face, T face = trustworthy face; VS = victim sensitivity; 95% CI = 95% confidence interval, *LL* = lower limit, *UL* = upper limit.

* Significant correlation ($p < .05$, two-tailed).

⁺ Significant correlation ($p < .05$, one-tailed).

Appendix A: Results of the competition conditions

Table A1

Means and Standard Deviations of Eye Movement Variables for Relevant Words

Competition Conditions

Word	First fixation duration [ms]	Dwell time [ms]	Fixation count	Destination of the first saccade [%]
U face				
U word	260 (43)	487 (76)	2.05 (0.36)	23.86 (5.27)
T word	259 (46)	481 (76)	2.03 (0.35)	25.39 (6.55)
T face				
U word	259 (51)	492 (82)	2.09 ^b (0.40)	25.54 (5.50)
T word	259 (46)	484 (79)	2.03 ^b (0.35)	24.70 (5.52)

Note. U face = untrustworthy face, T face = trustworthy face, U word = untrustworthy word, T word = trustworthy word. Values reported for “destination of the first saccade” represent the percentage of trials in which the word was fixated first. Standard deviations are shown in parentheses.

^b Significant difference between trustworthy and untrustworthy words in one face condition ($p < .05$, two-tailed).

Table A2*Correlations of Attentional Bias Scores with Victim Sensitivity**Competition Conditions*

Bias score	VS	<i>p</i>	95% CI	
			<i>LL</i>	<i>UL</i>
U face				
First fixation duration	.05	.689	-.189	.283
Dwell time	.19	.111	-.049	.408
Fixation count	.21 ⁺	.079	-.028	.426
Destination of the first saccade	-.06	.623	-.293	.179
T face				
First fixation duration	-.02	.878	-.255	.218
Dwell time	-.11	.370	-.338	.130
Fixation count	-.15	.212	-.373	.090
Destination of the first saccade	.04	.764	-.199	.274

Note. U face = untrustworthy face, T face = trustworthy face; VS = victim sensitivity; 95% CI = 95% confidence interval, *LL* = lower limit, *UL* = upper limit.

⁺ Significant correlation ($p < .05$, one-tailed).

Appendix B: Results of the multilevel analysis

Table B1

Mixed Model Results

	First fixation duration				Dwell time				Fixation count			
	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p
Random effects												
Residual	19,098.23	231.22	82.60	<.01	43,339.79	523.21	82.83	<.01	0.78	0.01	82.83	<.01
Intercept	1,719.45	309.11	5.56	<.01	4,409.43	787.65	5.60	<.01	0.11	0.02	5.68	<.01
Fixed effects	Estimate	SE	t	p	Estimate	SE	t	p	Estimate	SE	t	p
Intercept	262.56	5.13	51.19	<.01	482.04	8.19	58.87	<.01	2.03	0.04	49.21	<.01
Face	-0.46	2.36	-0.19	.85	-0.35	3.55	-0.10	.92	-0.01	0.02	-0.63	.53
Word	-4.33	2.36	-1.83	.07	15.19	3.55	4.28	<.01	0.10	0.02	6.92	<.01
ZVS	2.28	5.17	0.44	.66	-15.95	8.25	-1.93	.06	-0.08	0.04	-1.85	.07
Face x Word	0.49	4.72	0.10	.92	-11.38	7.09	-1.61	.11	-0.03	0.03	-1.03	.30
ZVS x Face	-4.91	2.38	-2.06	.04	-4.67	3.57	-1.31	.19	0.01	0.02	0.57	.57
ZVS x Word	2.74	2.38	1.15	.25	-4.10	3.57	-1.15	.25	-0.05	0.02	-3.07	<.01
ZVS x Face x Word	-6.85	4.76	-1.44	.15	-14.90	7.15	-2.09	.04	<0.01	0.03	0.07	.95
Model summary	Parameters	-2 Log-Likelihood			Parameters	-2 Log-Likelihood			Parameters	-2 Log-Likelihood		
Information criteria	10	174,305.14			10	186,605.96			10	36,025.44		

Note. Random effects: random intercepts were modeled for participants to account for the nested data structure. Fixed effects: two level-1 predictors (Face and Word) were included for face type (untrustworthy = -0.5, trustworthy = +0.5) and word type (untrustworthy = -0.5, trustworthy = +0.5). At level 2, victim sensitivity (ZVS; z-standardized) as well as the respective interaction terms were entered.

Table B2*Mixed Model Results Separately for Face Conditions*

	First fixation duration				Dwell time				Fixation count			
	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p
	U face											
Random effects	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p
Residual	20,069.69	344.52	58.25	<.01	42,941.83	735.04	58.42	<.01	0.79	0.01	58.42	<.01
Intercept (subject)	1,714.04	326.17	5.26	<.01	4,474.48	835.03	5.36	<.01	0.11	0.02	5.48	<.01
Fixed effects	Estimate	SE	t	p	Estimate	SE	t	p	Estimate	SE	t	p
Intercept	262.82	5.27	49.88	<.01	482.23	8.43	57.20	<.01	2.04	0.04	49.11	<.01
Word	-4.57	3.42	-1.34	.18	20.89	4.99	4.19	<.01	0.12	0.02	5.60	<.01
ZVS	4.74	5.31	0.89	.38	-13.60	8.49	-1.60	.11	-0.08	0.04	-1.94	.06
ZVS x Word	6.15	3.45	1.78	.08	3.36	5.03	0.67	.50	-0.05	0.02	-2.20	.03
	T face											
Random effects	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p	Variance	SE	Wald Z	p
Residual	18,097.46	310.62	58.26	<.01	43,817.06	749.91	58.43	<.01	0.77	0.01	58.43	<.01
Intercept (subject)	1,754.69	329.84	5.32	<.01	4,262.82	800.43	5.33	<.01	0.12	0.02	5.51	<.01
Fixed effects	Estimate	SE	t	p	Estimate	SE	t	p	Estimate	SE	t	p
Intercept	262.34	5.30	49.52	<.01	481.86	8.25	58.38	<.01	2.03	0.04	47.58	<.01
Word	-4.08	3.25	-1.25	.21	9.49	5.04	1.88	.06	0.09	0.02	4.19	<.01
ZVS	-0.19	5.34	-0.04	.97	-18.29	8.32	-2.20	.03	-0.07	0.04	-1.69	.10
ZVS x Word	-0.71	3.27	-0.22	.83	-11.55	5.08	-2.27	.02	-0.05	0.02	-2.14	.03

Note. U face = untrustworthy face, T face = trustworthy face. Random effects: random intercepts were modeled for participants to account for the nested data structure. Fixed effects: one level-1 predictor (Word) was included for word type (untrustworthy = -0.5, trustworthy = +0.5). At level 2, victim sensitivity (ZVS; z-standardized) as well as the respective interaction term was entered.

RELATIVE CONTRIBUTIONS TO THE PUBLICATIONS

Manuscript A: *Measuring justice sensitivity in patients with borderline personality disorder: Do we really measure the same?* Manuscript submitted for publication in *Assessment*.

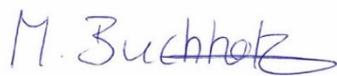
Buchholz, M.:	85%
Neukel, C.:	2,5%
Steinmann, S.:	2,5%
Lis, S.:	5%
Gollwitzer, M.:	5%

Manuscript B: *Taking back control: Findings on the cognitive, behavioral, and motivational consequences of victim sensitivity.* Manuscript submitted for publication in *Social Justice Research*.

Buchholz, M.:	65%
Magraw-Mickelson, Z.:	5%
Stolz, C.:	10%
Süssenbach, P.:	5%
Gollwitzer, M.:	15%

Manuscript C: *Victim sensitivity predicts attention allocation towards violations of untrustworthiness expectancies.* Manuscript submitted for publication in *Social Justice Research*.

Buchholz, M.:	75%
Bergmann, N.:	20%
Schubö, A.:	2,5%
Gollwitzer, M.:	2,5%



Signature PhD student (Merle Buchholz)



Signature PhD supervisor (Mario Gollwitzer)

CURRICULUM VITAE

Pages 185 to 187 contain personal data and are therefore not part of the online publication.

Pages 185 to 187 contain personal data and are therefore not part of the online publication.

Pages 185 to 187 contain personal data and are therefore not part of the online publication.

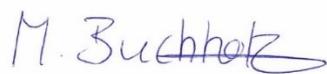
EIGENSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich meine Dissertation

Cause and Effect in Victim Sensitivity: Analyses of Associated Social-Cognitive Processes

selbstständig und ohne unerlaubte Hilfe angefertigt habe und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg an der Lahn, 15. Januar 2021,



Merle Buchholz