Weiss, A., Burgmer, P., & Mussweiler, T. (2018). Two-faced morality: Distrust promotes divergent moral standards for the self versus others. *Personality and Social Psychology Bulletin*, *44*, 1712-1724. doi:10.1177/0146167218775693

accepted version before copy-editing

Two-faced morality:
Distrust promotes divergent moral standards for the self versus others

Alexa Weiss[1], Pascal Burgmer[1], & Thomas Mussweiler[1,2]

[1]*Social Cognition Center Cologne, University of Cologne, Germany*

[2]*London Business School, United Kingdom*

Word count: 10,129

Correspondence concerning this article should be sent to Alexa Weiss, University of Cologne, Department Psychologie, Richard-Strauss-Str. 2, 50931 Koeln, Germany, alexa.weiss@uni-koeln.de.

Acknowledgements

We would like to thank the members of the Social Cognition Center Cologne for fruitful discussions of this research. Furthermore, we would like to thank two anonymous revievers for their helpful comments. This research was funded by the German Research Foundation (DFG) through a Gottfried Wilhelm Leibniz Prize (MU 1500/5-1) awarded to Thomas Mussweiler and a research grant within the research unit Design and Behavior awarded to Thomas Mussweiler und Axel Ockenfels (MU 1500/7-2). A. Weiss and P. Burgmer contributed equally to this work.

Abstract

People do not trust hypocrites, because they preach water, but drink wine. The current research shows that, ironically, when we distrust, we become moral hypocrites ourselves. We argue that experiencing distrust alerts us to the possibility that others may intent to exploit us, and that such looming exploitation differentially affects moral standards for the self versus others. Four studies (N = 1,225) examined this possibility and its underlying motivational dynamic. Study 1 established a relationship between dispositional distrust and flexible, self-serving moral cognition. In Studies 2 and 3, participants experiencing distrust (vs. trust) endorsed more lenient moral standards for themselves than for others. Study 4 explored the role of the motivation to avoid exploitation in these effects. Specifically, participants' dispositional victim sensitivity moderated the effect of distrust on hypocrisy. Together, these findings suggest that individuals who distrust and fear to be exploited show self-serving, and hence untrustworthy, moral cognition themselves.

Keywords: Distrust, Trust, Moral Hypocrisy, Moral Judgment, Exploitation Avoidance, Victim Sensitivity

From international affairs between countries to extramarital affairs in romantic relationships, hypocrisy—the endorsement of different moral standards for the self versus others—appears to us both prevalent and outrageous (e.g., Laurent, Clark, Walker, & Wiseman, 2014). In fact, moral hypocrites are evaluated more negatively, and specifically less trustworthy, than liars who straightforwardly claim to behave morally when they do not (Jordan, Sommers, Bloom, & Rand, 2017); by expressing strict moral standards for others, moral hypocrites send an implicit, but nevertheless convincing social signal about their own moral standards and hence future behavior—without paying the costs of abiding to these standards themselves (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997). The present research suggests that when people experience distrust, they may themselves become what they despise—moral hypocrites. In other words, for good reasons, we may distrust other hypocrites, but the experience of distrust may turn us into moral hypocrites ourselves: it makes us endorse more lenient moral standards for ourselves than for others. If this is the case, then distrust and moral hypocrisy may be crucial ingredients to a vicious cycle that corrodes interpersonal relationships: distrusting another's potentially benign intentions promotes double moral standards for one's own and another's behavior, which will in turn spur further distrust and hypocrisy.

Moral Hypocrisy

Morality is a fundamental dimension of human social perception and behavior (Goodwin, Piazza, & Rozin, 2014). We frequently commit, are targets of, and evaluate moral transgressions (Hofmann, Wisneski, Brandt, & Skitka, 2014). People use others' overt behavior to draw inferences about their moral character; seemingly harmless behaviors may have severe social consequences if observers conclude that these can only be committed by someone with a flawed character (Uhlmann, Pizarro, & Diermeier, 2015; see also Rom, Weiss, & Conway, 2017). Accordingly, people have a strong need for a moral reputation, and

are willing to incur great costs to avoid that negative information about them will be spread (Vonasch, Reynolds, Winegard, & Baumeister, 2017). Beyond reputational concerns, people have an internal need to regard themselves as moral persons (Aquino & Reed, 2002; Monin & Jordan, 2009); in fact, people view moral traits as most essential to the self (Strohminger & Nichols, 2014).

This strong motivation to uphold a moral self-image can give rise to motivated biases in people's moral evaluations (Ditto, Pizarro, & Tannenbaum, 2009; Klein & Epley, 2017). Indeed, people demonstrate a notable flexibility in their moral judgment and oftentimes process information in a manner that confirms their preferred conclusions about the morality or immorality of certain behaviors (Haidt, 2001; Shalvi, Gino, Barkan, & Ayal, 2015). Consequently, lenient judgments of one's own moral transgressions while holding others to strict moral standards may serve to see the self in a particularly positive light, thus maintaining one's moral self-image. Such double moral standards are hence indicative of moral hypocrisy as an inauthentic, self-serving morality (Monin & Merritt, 2012). Earlier studies have assessed hypocrisy by revealing inconsistencies between what people claim to find moral and their actual behavior (Batson et al., 1997). More recent studies have compared people's moral standards for themselves to the standards they hold other people to (Lammers, 2012; Lammers, Stapel, & Galinsky, 2010; Polman & Ruttan, 2012; Valdesolo & DeSteno, 2008). In these studies, participants evaluate the moral acceptability of identical transgressions either from their own or from another person's perspective. The resulting double moral standards are indicative of not practicing what one preaches to others (Monin & Merritt, 2012).

Importantly, hypocritical moral standards may not only help individuals to deal with their past transgressions, but also rationalize future immoral behavior. Specifically, generating justifications and mitigating information before committing a potential

transgression enables people to behave selfishly and feel moral nevertheless (Shalvi et al., 2015).

Conversely, harsh moral judgments of others' transgressions may prevent immoral behaviors committed by others: the expression of moral disapproval can act as a powerful deterrent and punishment (Keltner & Haidt, 1999). Beyond this social norm enforcing function of moral judgments, strong emotional reactions alert the self, thus preparing oneself for and preventing detrimental consequences of others' transgressions (Haidt, 2003). In sum, double moral standards are functional in pursuing one's self-interest and protecting it from others' behaviors (Monin & Merritt, 2012). Protecting one's self-interest, in turn, is particularly important when we suspect that others might betray and harm us—when we distrust.

Distrust and Moral Hypocrisy

When trusting others, we expect that they will act in our best interest and have good intentions. Distrust, on the other hand, alerts us to others' potentially malevolent intentions. Positive expectations of another's intentions and behavior entail the acceptance of vulnerability (Rousseau, Sitkin, Burt, & Camerer, 1998). In contrast, then, doubting another's intentions should entail a state of mind characterized by attempts to minimize vulnerability, and an enhanced motivation to avoid exploitation by one's interaction partners. Consequently, when conceiving of the risk that someone might betray their interests, individuals should become particularly motivated to protect their own self-interest and to restrain their counterpart, thereby counteracting any potential vulnerability (Lewicki, McAllister, & Bies, 1998). We therefore contend that the motivation to avoid exploitation associated with distrust will affect moral judgments of own versus others' (potential) transgressions in diverging ways, and hence promote moral hypocrisy. Previous research, particularly in the domain of social dilemmas, indeed found that distrusting individuals

behave less cooperatively (De Cremer, Snyder, & Dewitte, 2001; McCabe, Rigdon, & Smith, 2003). Arguably, individuals who fear to end up as the "sucker" are predominantly focused on their self-interest and motivated to behave selfishly in a preemptive manner (Lewicki et al., 1998). In consequence, they may search for mitigating information to justify self-interested transgressions, leading to lenient moral standards for the self.

At the same time, distrusting individuals may attend to aggravating information or applicable moral rules for others, motivated to affirm and enforce moral standards that constrain others' behaviors. In particular, harsh moral judgments of other's potential immoral behaviors may serve to nip their exploitative attempts in the bud, sending a signal to potential transgressors. In addition to this social function, as outlined above, moral emotions such as anger serve an important alerting and protective function in the face of others' harmful behaviors (Haidt, 2003). Moreover, under distrust, people may question their counterpart's intentions for ambiguous behaviors (Hilton, Fein, & Miller, 1993). Recent research found that people predict themselves to be less capable of and to feel worse after behaving immorally, compared to other persons. This self-righteousness effect stems from an asymmetrical perception of intentions: people tend to perceive their own intentions as more ethical—even when they behave unethically (Klein & Epley, 2017). Importantly, under distrust, this tendency should be even increased, promoting hypocrisy in moral judgments of transgressions.

In line with the above reasoning, the few previous empirical investigations of the antecedents of moral hypocrisy suggest that it is increased by a motivational focus on one's own interests at the expense of the interests of other parties. For example, social power increases double moral standards (Lammers, Stapel & Galinsky, 2010). Power has been linked to both a sense of entitlement to rewards (Keltner, Gruenfeld, & Anderson, 2003) and enhanced goal-pursuit (Burgmer & Englich, 2012), and decreased concern for other's

perspectives (Galinsky, Magee, Inesi, & Gruenfeld, 2006). In contrast, hypocrisy is attenuated by feelings of guilt (Polman & Ruttan, 2012)—an affective reaction to having harmed another individual, for example by a selfish act.

Furthermore, the cognitive underpinnings of experiencing distrust may contribute to double moral standards. Previous research found that distrust cognitively tunes people to consider aspects of situations that may be different than they appear and to entertain alternative interpretations (Kleiman, Sher, Elster, & Mayo, 2015; Schul, Mayo, & Burnstein, 2004; see Mayo, 2015, for an overview). In particular, distrust enhances cognitive flexibility and creativity (Mayer & Mussweiler, 2011). Cognitive flexibility, in turn, facilitates self-serving justifications and hence immoral behaviors such as cheating (Gino & Ariely, 2012). An abstract, and hence more flexible thinking style has further been shown to increase moral hypocrisy (Lammers, 2012). In sum, a suspicious mind appears well prepared for moral hypocrisy already on a basic level of information processing.

On a motivational level, previous theorizing, and some empirical research particularly in the domain of social dilemmas, has invoked the notion of a motivation to avoid exploitation (Carpenter & Dolifka, 2017; De Cremer, 1999; Effron & Miller, 2011; Yamagishi et al., 2017). It has been argued that individual differences in "sugrophobia," the "fear of being suckered," reflect affective reactions to being exploited (Vohs, Baumeister, & Chin, 2007). Similarly, victim sensitivity describes interindividual differences in reactions to being the victim of injustice and hence a motivation to avoid being exploited (Baumert & Schmitt, 2016; Schmitt, Gollwitzer, Maes, & Arbach, 2005; Gollwitzer, Süssenbach, & Hannuschke, 2015). Victim sensitivity also involves an enhanced sensitivity to untrustworthiness cues (Gollwitzer, Rothmund, Pfeifer, & Ensenbach, 2009). According to the Sensitivity of Mean Intentions (SeMI) model, such cues readily elicit a "suspicious mindset" in highly victim-sensitive individuals, which entails legitimizing cognitions for own

immoral behaviors and hostile interpretations of other's behaviors (Gollwitzer, Rothmund, & Süssenbach, 2013).

Indeed, a number of empirical studies support the hypothesis that when distrust cues are present, victim-sensitive individuals show higher levels of uncooperative or selfish behavior, arguably in an attempt to protect the self and secure their share (Gollwitzer et al., 2013). However, while most studies have looked at behavioral variables, particularly in social dilemmas, only few have assessed victim-sensitive individuals' evaluations of transgressions (Gerlach, Allemand, Agroskin, & Denissen, 2012; Gollwitzer, Schmitt, Schalke, Maes, & Baer, 2005). Yet, if dispositional victim sensitivity amplifies reactions to untrustworthiness out of self-protective concerns, it should also exacerbate the hypocrisy-inducing effects of distrust.

The Present Research

The present research investigates how the experience of distrust (vs. trust) shapes the propensity for moral hypocrisy. We hypothesize that distrustful individuals will exhibit greater leniency for their own compared to other persons' moral transgressions. This effect may be partially driven by distrustful people's motivation to avoid exploitation by others and should thus be moderated by dispositional victim sensitivity. We examined this possibility in four studies. In Study 1, we explored dispositional associations between social distrust and self-serving bending of moral rules (moral flexibility) as an indicator of hypocritical tendencies. In Studies 2-4, we experimentally investigated how the experiences of distrust (vs. trust) affect the perceived acceptability of various daily-life moral transgressions either from participants' own or third persons' perspectives. In Study 2, we employed an imagination task that manipulated participants' distrust (trust) in a hypothetical interaction with a target person. In Study 3, an episodic recall task was used to orthogonally manipulate distrust (trust) while measuring moral judgments of various independent scenarios. We

expected distrusting participants to be more lenient in judging their own transgressive behavior versus the behavior of others, even if these others are actually unrelated to the source of their distrust. Study 4 explored the role of victim sensitivity, investigating whether participants' dispositional motivation to avoid exploitation would amplify the effect of distrust on moral hypocrisy.[1] Specifically, when experiencing distrust, high victim-sensitive individuals should be particularly inclined to entertain diverging moral standards for themselves versus others. In contrast, low victim-sensitive individuals should be less sensitive to such untrustworthiness cues.

Study 1

Study 1 was designed as an initial examination of whether distrust is related to flexible and self-serving moral judgments. To do so, we followed a correlational approach: we used a questionnaire to assess dispositional tendencies to distrust other people (Yamagishi, 1988). In addition, participants completed a scale previously used to measure moral disengagement and flexibility about moral rules if it serves one's self-interest as a proxy for dispositional moral hypocrisy (Cameron & Payne, 2012; Shu, Gino, & Bazerman, 2011). Such self-serving, opportunistic adjustment of moral judgments has previously been described as a central characteristic of hypocrisy (Monin & Merritt, 2012). We expected participants higher in dispositional distrust to score higher on this moral flexibility scale.

Method

Participants and design. Sample size was determined expecting at least $r = .20$, which requires a minimum sample of approximately $N = 240$. We obtained data from 246 U.S. American adults (148 males, 98 females, $M_{age} = 34.40$, $SD_{age} = 11.53$) via Amazon Mechanical Turk (MTurk), who received $0.30. Six additional participants were excluded for failing at an attention-check item.

[1] For all studies, we report all relevant measures, all conditions, all data exclusions, and we provide a rationale for how sample sizes were determined. Materials for all studies can be found in the Supplementary Online Materials (SOM).

Materials and procedure. Participants filled out a dispositional distrust questionnaire consisting of eight items (e.g., "One should not trust others until one knows them well"; Yamagishi, 1988) and a six-item scale assessing moral flexibility (Shu et al., 2011) on scales from 1 (= *strongly disagree*) to 7 (= *strongly agree*). The moral flexibility scale comprises items that measure the perceived acceptability of moral rule violations if these serve one's interest (e.g., "Rules should be flexible enough to be adapted to different situations"). Order of measures was counterbalanced between participants. We averaged both scales to indices of dispositional distrust (Cronbach's $\alpha = .87$) and moral flexibility ($\alpha = .78$). Following the first measure, participants were prompted to move a slider to a certain scale-point as an attention-check item.

Results and Discussion

Confirming expectations, participants' dispositional distrust ($M = 4.30$, $SD = 1.07$) was significantly associated with their moral flexibility ($M = 3.39$, $SD = 1.02$), $r(246) = .326$, $p < .001$, 95% CI [.199, .439].[2] Individuals who were more distrusting of other people expressed greater acceptance of bending moral rules for self-interested reasons.

This finding provides initial support for the idea that a distrusting state of mind may promote a self-interested shifting of moral standards, which is indicative of moral hypocrisy (Monin & Merritt, 2012). Yet, this result is only correlational and therefore does not speak to causality. What is more, it remains quite possible that distrustful individuals are simply more accepting of moral transgressions in general, irrespective of the agent. However, we reasoned that distrusting individuals should be selectively more lenient about their own moral transgressions, but not others' immoral behaviors—a dissociative pattern of self- versus other-related moral judgments.

Study 2

[2] Results of Studies 1 and 4 are based on bootstrapping with 5,000 samples.

Study 2 employed an experimental design to investigate how experiences of distrust (vs. trust) shape moral hypocrisy. Participants experienced distrust (vs. trust) towards an imagined target person (a co-worker) and subsequently judged different moral transgressions that involved both the participant and the co-worker. These scenarios pitted the self-interest of one of them against the other's interest or a general moral rule or mutual agreement. Depending on condition, either the participant or the co-worker were presented as the (potential) enactor of the behaviors and target of judgment in these scenarios. We predicted that distrustful participants would make more lenient moral judgments for transgressive behaviors committed by themselves compared to when committed by their counterpart. For participants experiencing trust, such moral hypocrisy should not emerge.

Method

Participants and design. We conducted an a priori power analysis with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Expecting a small-to-medium interaction effect, and requiring 80% power, we aimed at collecting approximately $N = 270$ participants. We recruited 271 U.S. American adults (115 males, 156 females, $M_{age} = 34.57$, $SD_{age} = 12.78$) via MTurk, who received $0.40. Participants were randomly assigned to one of four conditions of a 2 (experience: distrust vs. trust) × 2 (target: self vs. co-worker) between-subjects design. In total, eleven additional participants were excluded from analyses because they indicated having answered at least one item in a random fashion or purposely wrong ($n = 5$) or having participated multiple times in the same survey ($n = 5$), or because they could or did not provide any sensible response to the imagination task ($n = 3$). In Studies 2-4, multiple exclusion criteria per participant were possible.

Materials and procedure. Participants were asked to complete an imagination task. First, they were asked to imagine that they had a new job and that they had been assigned to work in a team towards a goal of their company. Subsequently, they learned that a new co-

worker also joined the team. In the *distrust* condition, the co-worker's past behavior was described as unreliable, and participants were told that his or her intentions seemed questionable, so that they could not trust that person. Contrary, in the *trust* condition, this co-worker's past behavior was described as conveying reliability and benign intentions so that they could trust this person (see SOM). In order to strengthen the effect of the manipulation, participants were then asked to write down some of their thoughts about the imagined situation (i.e., how they would think, feel, and act).

Subsequently, participants imagined four additional situations that might happen at that workplace. They saw four moral scenarios in a fixed random order that each described a situation where either they themselves (*self* condition) or their co-worker (*co-worker* condition) considered breaking a rule for self-interested reasons. For instance, participants in the co-worker condition read that due to time pressure, the co-worker considered to secretly delegate an important task to a less qualified intern despite a contrary agreement (see SOM). For each of the scenarios, participants rated the degree of acceptability of the described behavior on a 7-point scale ranging from *not at all acceptable* to *completely acceptable*. We averaged ratings across scenarios to form an acceptability index ($\alpha = .66$) with higher values indicating higher moral leniency. Subsequently, participants recalled the imagination task and indicated whether they thought they could trust the described co-worker on a 7-point scale ranging from *not at all* to *absolutely*. This measure served as a manipulation check.

Results and Discussion

Our manipulation of distrust (vs. trust) was successful. A 2 (experience: distrust vs. trust) × 2 (target: self vs. co-worker) ANOVA on the manipulation-check item confirmed that participants experienced more distrust towards the co-worker in the distrust condition ($M =$

1.43, *SD* = 0.90) compared to the trust condition (*M* = 6.08, *SD* = 1.41), *F*(1, 267) = 1045.42, $p < .001$, $\eta_p^2 = .797$, 90% CI [.764; .821][3]. No other effects emerged (*F*s ≤ 0.18).

Consistent with predictions, in the distrust condition, participants who judged moral transgressions from their own perspective (*M* = 2.43, *SD* = 1.00) judged these more leniently compared to participants who imagined their co-worker as the transgressor (*M* = 1.80, *SD* = 0.79), *t*(267) = 3.82, $p < .001$, Cohen's *d* = 0.65, 95% CI [0.31; 0.99]. Under distrust, participants hence exhibited moral hypocrisy in their judgments (Figure 1). In contrast, in the trust condition, moral judgments of the behaviors did not differ as a function of target condition (*M* = 2.15, *SD* = 0.89, self condition; *M* = 2.29, *SD* = 1.11, co-worker condition), *t*(267) = -0.89, *p* = .375. A 2 (experience: distrust vs. trust) × 2 (target: self vs. co-worker) ANOVA revealed the predicted interaction effect, *F*(1, 267) = 11.04, *p* = .001, $\eta_p^2 = .040$, 90% CI [.010; .084]. In addition, a significant main effect of target emerged, *F*(1, 267) = 4.26, *p* = .040, $\eta_p^2 = .016$, 90% CI [.000; .049]. Moral judgments in the self-condition (*M* = 2.29, *SD* = 0.96) were overall more lenient than in the co-worker condition (*M* = 2.05, *SD* = 0.99), *t*(269) = 2.02, *p* = .045, *d* = 0.25, 95% CI [0.01; 0.48]. There was no main effect of experience, *F*(1, 267) = 0.84, *p* = .359.

These findings extend the correlational data of Study 1 in multiple ways. First, they indicate that distrust has a causal effect on moral judgment. Second, the effect of distrust on moral judgment critically depends on the target of judgment: under distrust, participants gave more lenient judgments of the same transgressions when these were presented from their own than from their co-workers perspective, whereas under trust, participants judged their own and their co-worker's moral transgressions equally. This indicates that moral judgment is not generally less rule-based and more lenient under distrust. Instead, distrusting individuals

[3] For one-sided tests (*F*-tests), 90% confidence intervals are reported (Lakens, 2014; Steiger, 2004).

seem to interpret moral rules and situations according to their self-interest, resulting in double moral standards.

Study 2 suggests that individuals who distrust another person show moral hypocrisy with respect to this interaction partner. However, previous findings on the antecedents of moral hypocrisy, such as social power (Lammers et al., 2010), suggest that the effects of a focus on protecting one's own interest and avoiding exploitation associated with distrust may even extend beyond a particular situation and relationship partner. Indeed, distrusting individuals may sometimes even react uncooperatively towards targets completely unrelated to the previous untrustworthy situation or interaction partner (Rothmund, Gollwitzer, & Klimmt, 2011). In addition, the experience of distrust may lead to enhanced attributions of malevolent intentions to others (Fein et al., 1993). Moreover, on a cognitive level, cues of untrustworthiness may elicit a distrust mindset that generalizes to subsequent situations (Mayo, 2015). Hence, extending the previous study, we designed Study 3 to examine whether the effect of distrust generalizes to moral judgments with respect to persons completely unrelated to the source of distrust.

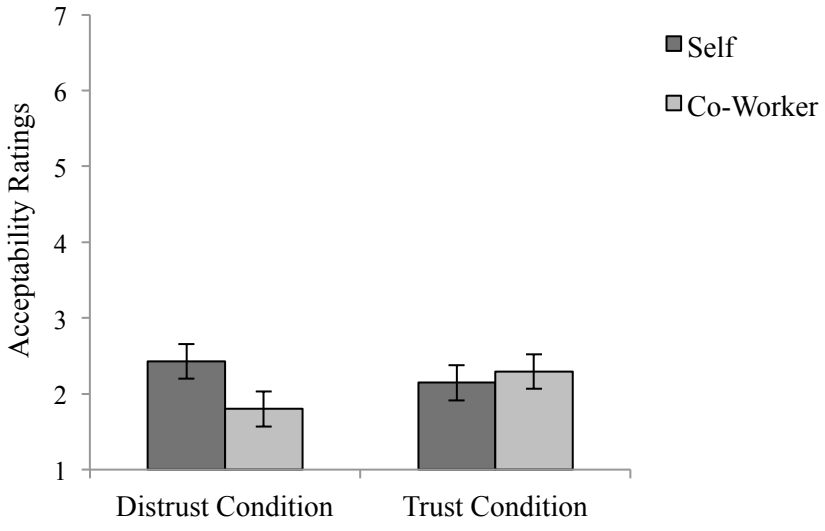

Figure 1: Moral judgments as a function of experience and target conditions (Study 2). Participants' mean acceptability ratings of the imagined moral transgressions committed either by themselves or committed by a co-worker who had been described as either untrustworthy or trustworthy. Higher values indicate higher perceived acceptability (scale 1-7). Error bars reflect 95% confidence intervals.

Study 3

In this study, we used an episodic recall task as manipulation of experiences of distrust versus trust (Kleiman et al., 2015). We also designed new moral scenarios that cover a wide range of daily-life situations. Accordingly, these scenarios did not involve the same social target and context that previously prompted the experience of distrust or trust, thus allowing us to test a potential spill-over effect of distrust on moral hypocrisy. We again expected a pattern of increased moral hypocrisy under distrust (i.e., greater leniency towards own vs. others' moral transgressions) compared to trust.

Method

Participants and design. Based on the effect size obtained in Study 2, the current study required a sample size of approximately 200 individuals. Taking into account potential exclusions and—due to the predicted spill-over effect—a potentially weaker effect than in Study 2, we obtained data from 217 U.S. American adults (89 males, 128 females, M_{age} = 34.24, SD_{age} = 13.10) via MTurk, who received $0.45. Another 26 additional participants were excluded from analyses because they indicated random or wrong responding (n = 2), did not provide a sensible answer to the recall task (n = 2) or failed a manipulation-check item at the end of the study prompting them to identify the theme of the recall task (i.e., distrust vs. trust; n = 22). Participants were randomly assigned to one of four conditions in a 2 (experience: distrust vs. trust) × 2 (target: self vs. other) between-subjects design.

Materials and procedure. Presented as a task on autobiographical experiences, participants in the *distrust* (*trust*) condition first wrote about a time in their life when they distrusted (trusted) another person. In both conditions, participants were instructed to recall a

situation where their distrust (trust) was justified in retrospect. Moreover, they were asked to keep the situation in mind because they would later be asked to proceed with this task (see SOM; see Kleiman et al., 2015, for a similar procedure).

Next, as a task on situation perception, participants read six moral scenarios structurally similar to those in Study 2 in a fixed random order. These scenarios described different daily-life moral transgressions, for example keeping too much change received from a cashier or submitting an unjustified warranty claim after accidentally damaging one's cell phone (see SOM). Two scenarios were adapted from previous research (Lammers, 2012; Lammers et al., 2010). For each scenario, participants rated how acceptable they would find it for themselves (*self* condition) or another unspecified person (*other* condition) to engage in the described behaviors. We averaged ratings across scenarios to form an acceptability index ($\alpha = .78$).

Results and Discussion

Consistent with expectations, in the distrust condition, participants who judged their own transgressions ($M = 3.55$, $SD = 1.35$) were more lenient than those judging others' transgressions ($M = 2.99$, $SD = 1.44$), $t(213) = 2.22$, $p = .028$, $d = 0.44$, 95% CI [0.05; 0.84]. Such a pattern of moral hypocrisy did not emerge in the trust condition ($M = 3.16$, $SD = 1.12$, self condition; $M = 3.49$, $SD = 1.16$, other condition), $t(213) = -1.42$, $p = .158$ (Figure 2). In addition, a 2 (experience: distrust vs. trust) × 2 (target: self vs. other) ANOVA on participants' acceptability ratings revealed the predicted interaction effect, $F(1, 213) = 6.70$, $p = .010$, $\eta_p^2 = .031$, 90% CI [.004; .077]. No other effects emerged (all Fs ≤ 0.44).

Study 3 hence replicates the results of Study 2 with a different distrust manipulation and a new set of various daily-life scenarios. Ruling out potential demand effects, the current findings indicate that the experience of distrust (vs. trust) promotes moral hypocrisy even with respect to transgressions that are unrelated to the incidental source of distrust.

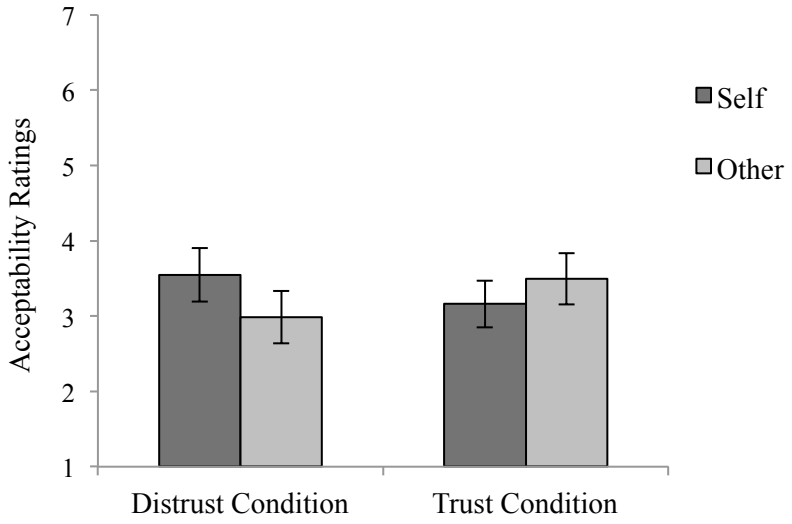

Figure 2. Moral judgments as a function of experience and target conditions (Study 3). Participants' mean acceptability ratings of the imagined moral transgressions either committed by themselves or committed by another person as a function of distrust (vs. trust). Higher values indicate higher perceived acceptability (scale 1-7). Error bars reflect 95% confidence intervals.

Thus, Studies 1-3 indicate that dispositional, situational, and incidental distrust (vs. trust) increase moral flexibility and moral hypocrisy. Study 4 explored a motivational dynamic that may contribute to this effect. Specifically, we examined whether dispositional victim sensitivity—characterized by a high motivation to avoid exploitation—would moderate hypocritical tendencies in response to distrust.

Study 4

Study 4 was closely designed after Study 2. We therefore expected to replicate the interaction between experiences of distrust and target of moral judgment found in previous experiments, such that distrustful (vs. trustful) participants would again show a stronger moral hypocrisy effect. In addition, Study 4 investigated whether trait differences in victim sensitivity moderate the effect of distrust on moral standards for the self versus others. Based on the assumption that distrusting another person increases the motivation to protect oneself from being exploited and to protect one's self-interest, we predicted that individuals with

high victim sensitivity—that is, strong dispositional self-protective concerns (Schmitt et al., 2005)—would show stronger hypocrisy under conditions of distrust.

Method

Study 4 comprised two different data collection phases on MTurk. In the intake survey, we invited a large number of participants to complete a short personality survey including a measure of trait victim sensitivity (Schmitt et al., 2005). These participants were later invited to participate in the main study, which closely followed the design and procedure of Study 2. Participants completed the main study between approximately one to five weeks after the intake survey. This temporal dissociation between measurements of victim sensitivity as dispositional moderator and moral judgments as focal outcome variable provided a particularly strong test of our moderation hypothesis, preventing potential demand effects.

Participants and design. A total of 900 participants completed a short personality survey, and received $0.30 as compensation. At the end of the survey, participants were asked to provide their MTurk Worker-ID, a unique account identifier, if they agreed to be invited to a subsequent study. The main study was framed as a study on situation perception and accessible to all participants who had indicated their Worker-ID at the end of the intake survey. Participants received $0.45 upon completion. We aimed at collecting a minimum sample of $n = 100$ participants per between condition. However, anticipating a considerable drop out, we allowed all participants from the intake survey to participate in the main study. Data from six participants in the main study could not be matched with the data set from the intake survey. Furthermore, some participants who provided data to both the intake survey and main study were excluded from analyses because they indicated to have participated multiple times in the intake survey ($n = 4$) or in the main study ($n = 10$), or admitted to random or purposefully wrong responding in the intake survey ($n = 6$) or the main study ($n =$

2). Exclusions were also based on a failed attention check in the main study (n = 14) or on not providing meaningful text input for the imagination task (n = 1). In addition, due to technical difficulties, some participants (n = 14) were able to participate twice with their identical account in two different MTurk batches, so that the second data set provided by the respective participant was excluded from analyses. Taking into account these exclusions, we obtained complete data from 491 participants (257 females, 232 males, 2 other, M_{age} = 36.97, SD_{age} = 11.92).

Materials and procedure. In the intake survey, participants completed the short version of the victim sensitivity scale (Schmitt et al., 2005). Sample items are "I ruminate for a long time when other people are treated better than me" or "I cannot easily bear it when others profit unilaterally from me". Participants rated their agreement with ten statements on six-point scales (1 = *not at all*; 6 = *exactly*; M = 3.94, SD = 1.09. α = .92).[4]

For the main study, we slightly adjusted the four scenarios from Study 2 in order to reduce a potential floor effect (see SOM). Specifically, in order to increase the moral wiggle room in the scenarios, the target's incentives and/or justifications for the immoral behavior in question were stressed more. In addition, we extended the moral judgment measure from one to four items per scenario and had participants indicate their judgments on nine-point scales (1 = *not at all*, 9 = *absolutely*). For example, the four items for the scenario "delegation" in the *self* condition read "I would probably have good reasons to delegate the report to the intern", "It would reflect poorly on me if I delegated the report to the intern" (reversed), "The circumstances would justify that I delegate the report to the intern", and "It would be a questionable choice on my part if I delegated the report to the intern" (reversed). Mean ratings again reflected more lenient moral judgments (α = .66).

Results and Discussion

[4] We also assessed and controlled for participants' justice sensitivity from an observer's perspective (see SOM). Subsequently, participants completed personality measures not relevant to the present studies.

Effect of distrust on moral hypocrisy. We first submitted the data from the main study to a 2 (experience: distrust vs. trust) × 2 (target: self vs. co-worker) ANOVA. Consistent with expectations and replicating the results of Studies 2 and 3, in the distrust condition, participants who judged their own transgressions (M = 3.77, SD = 1.26) exhibited greater leniency compared to those who judged their co-worker's transgressions (M = 2.70, SD = 1.09), $t(487) = 6.83$, $p < .001$, d = 0.86, 95% CI [0.61; 1.12]. In contrast, participants who experienced trust judged transgressions in the self condition less leniently than in the co-worker condition, thus showing the reversed pattern (M = 3.36, SD = 1.29, self condition; M = 3.74, SD = 1.30, co-worker condition), $t(487) = -2.39$, $p = .017$, d = -0.31, 95% CI [-0.56; -0.05]. Accordingly, participants' acceptability ratings revealed the predicted interaction effect, $F(1, 487) = 42.09$, $p < .001$, $\eta_p^2 = .080$, 90% CI [.045; .120]. Significant main effects of experience, $F(1, 487) = 8.11$, $p = .005$, $\eta_p^2 = .016$, 90% CI [.003; .040], and target, $F(1, 487) = 9.47$, $p = .002$, $\eta_p^2 = .019$, 90% CI [.004; .044], emerged as well. Participants were overall more lenient in the self (M = 3.57, SD = 1.29) than in the co-worker (M = 3.22, SD = 1.30) condition, $t(489) = 3.01$, $p = .003$, d = 0.27, 95% CI [0.09; 0.45]. They also gave less lenient judgments under distrust (M = 3.24, SD = 1.29) compared to trust (M = 3.56, SD = 1.30), $t(489) = -2.71$, $p = .007$, d = -0.24, 95% CI [-0.42; -0.07].

Victim sensitivity as moderator. Subsequently, moral acceptability ratings were regressed on target (0 = self, 1 = co-worker), experience (0 = trust, 1 = distrust), victim sensitivity, and their interaction terms. The predicted three-way interaction between target, experience, and victim sensitivity was significant, b = -0.516, SE = 0.203, $t(483) = -2.54$, p = .011, 95% CI [-0.915; -0.116], confirming that victim sensitivity moderated the effect of distrust on moral hypocrisy (for complete regression results, see SOM).

Specifically, in the distrust condition (Fig. 3, left panel), a significant target × victim sensitivity interaction emerged, b = -0.489, SE = 0.137, $t(483) = -3.56$, $p < .001$, 95% CI [-

0.759; -0.219]. In line with predictions, when experiencing distrust, participants' moral hypocrisy inclination was affected by their dispositional victim sensitivity. Analyses revealed that participants' victim sensitivity predicted greater leniency for moral judgments of their own transgressions, $b = 0.316$, $SE = 0.098$, $t(483) = 3.24$, $p = .001$, 95% CI [0.124; 0.507]. In contrast, there was a marginally significant effect of victim sensitivity in the opposite direction on judgments of the co-worker's transgressions, $b = -0.173$, $SE = 0.097$, $t(483) = -1.79$, $p = .074$, 95% CI [-0.364; 0.017]. In other words, moral standards for themselves versus their co-worker diverged more strongly under distrust for participants with higher, compared to lower, levels of dispositional victim sensitivity.

In contrast, in the trust condition (Fig. 3, right panel), the target × victim sensitivity interaction was non-significant, $b = 0.027$, $SE = 0.150$, $t(483) = 0.18$, $p = .857$, 95% CI [-0.267; 0.321]. Higher victim sensitivity scores were associated, albeit non-significantly, with higher moral acceptability ratings for both the self, $b = 0.158$, $SE = 0.105$, $t(483) = 1.51$, $p = .131$, 95% CI [-0.047; 0.363], and the co-worker, $b = 0.185$ $SE = 0.107$, $t(483) = 1.72$, $p = .085$, 95% CI [-0.026; 0.396]. These results remain highly similar when justice sensitivity from an observers' perspective is included as a covariate (see SOM).

These results are in line with the notion that the motivation to avoid exploitation and to protect specifically one's self-interest contributes to the link between distrust and moral hypocrisy. When distrusting another person, participants exhibited double moral standards, and became more lenient with themselves than they were with their interaction partner. However, this hypocritical judgment tendency was particularly pronounced among high victim-sensitive participants, that is, individuals who are highly motivated to avoid exploitation. In contrast, when trusting another person, victim sensitivity did not predict moral hypocrisy. In line with our reasoning, these findings suggest that people may utilize

hypocritical moral judgments to counteract the threat of exploitation when perceiving the need to protect themselves under conditions of distrust.

Figure 3. Moral judgments as a function of target condition and victim sensitivity in the distrust (left panel) and trust (right panel) conditions (Study 4). Participants' mean moral judgments of the imagined moral transgressions either committed by themselves or committed by their co-worker at ±1 standard deviation of victim sensitivity in the distrust (left panel) and trust (right panel) conditions. Higher values indicate more lenient moral judgments (scale 1-9).

General Discussion

The present studies indicate that trust and distrust differentially affect flexible and self-serving moral judgments. Whereas individuals in a trustful state of mind appear to judge moral transgressions as more or less equally acceptable regardless of whether they are committed by themselves or another person, individuals experiencing distrust exhibit moral hypocrisy. They judge identical transgressions less harshly when committed by themselves rather than another individual. As an initial examination of our hypotheses, Study 1 employed a correlational design, showing that dispositional distrust is related to a flexible and self-serving rather than an absolute interpretation of moral rules. Distrustful individuals more strongly supported bending moral rules when it served their self-interest. Studies 2 and 3 provided experimental support for our hypotheses, applying different techniques to induce the experiences of distrust versus trust, that is, an imagination task and a recall task. In

addition, different sets of daily-life moral transgressions were used to assess moral judgments in these experiments. Study 3 additionally revealed that distrust can increase moral hypocrisy even for targets of judgment who are not related to the initial source of distrust. Study 4 replicated the findings from Studies 2 and 3. More importantly, it examined the role of the motivation to protect oneself from exploitation. We found that dispositional differences in victim sensitivity assessed in a separate survey moderated the effect of distrust on moral hypocrisy assessed weeks later. Individuals who are highly sensitive to being the victim of injustice reacted most strongly to a target's untrustworthiness versus trustworthiness. Importantly, under distrust, but not trust, victim sensitivity predicted judgments of one's own and another's moral transgressions in opposite directions. Taken together, distrust promotes moral hypocrisy particularly when people are motivated to avoid exploitation and protect their self-interest.

Methodological Contributions and Limitations

On a methodological level, the present research contributes to research on both trust and distrust and moral judgment. Specifically, we developed a new experimental manipulation to induce the experiences of trust and distrust (Studies 2 & 4). Further, the few previous experimental investigations of hypocrisy mostly relied on single scenarios. Here, we developed new scenario batteries covering various daily-life situations, hence providing a statistically more powerful measure of moral hypocrisy. In line with previous research (Lammers et al., 2010), we operationalized moral hypocrisy between subjects. Conceivably, within-subjects designs might disclose the research question to participants. Such direct comparisons between moral transgressions from one's own and others' perspectives could thus restrain motivated moral reasoning, because people value consistency and fairness, and believe in the objectivity of their moral positions (Epley & Caruso, 2004; Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011). In other words, rendering self-other differences salient to

participants in a within-subjects design might threaten their moral self-image. In addition, as noted above, appearing as a hypocrite is socially risky (Jordan et al., 2017). Within-subjects designs with greater temporal delay between self and other judgments, however, could provide an avenue for future research in this respect.

More generally, the present research solely relied on hypothetical scenarios, in line with most previous hypocrisy research. However, previous research revealed a potential disconnect between people's reactions to hypothetical as opposed to actual transgressions. For example, people may anticipate considerably stronger affective responses to racist acts (targeted at a member of a different ethnic group) compared to their actual affective and behavioral responses (Kawakami, Dunn, Karmali, & Dovidio, 2009). Similarly, judgments of others' actual behavior may be more strongly biased by whether and how it affects observers' own self-interest compared to judgments of merely imagined transgressions (Bocian & Wojciszke, 2014). Not only people's actual moral evaluations, but also their subsequent behaviors may be affected by various factors, including the avoidance of potential costs and risks (e.g., being punished). Hypocritical moral judgments may thus not directly translate into behavior. Nevertheless, previous studies in the domain of moral hypocrisy have found convergence between moral judgment and behavior, and hypocritical moral standards with respect to actual transgressions. For example, one study found that power increased actual cheating behavior in accordance with its effects on leniency in self-related moral judgments (Lammers et al., 2010, Study 1). Valdesolo and DeSteno (2008) compared judgments of participants' own actual selfish behavior in the laboratory to judgments of another participant's ostensible selfishness, exposing moral hypocrisy. The current limitations notwithstanding, these previous findings may thus suggest that the motivational dynamics involved in the present research can translate to moral reactions in real-life situations. In addition, as moral judgments can serve as justifications for future behavior (Shalvi et al.,

2015), hypocritical standards should affect the actual tendency to break rules, and to punish such behaviors in others.

Another limitation of the present research is that it focused on interpersonal relationships, looking at individual persons as both sources of distrust and agents of transgressions. Future research might thus investigate how the present findings translate to intergroup contexts. Findings that in an intergroup conflict, group members may focus more strongly on moral values of loyalty compared to fairness or harm, exemplify how group membership affects the biased justification of transgressions (Leidner & Castano, 2012). Interestingly, victim sensitivity can contribute to exploitative interpretations of intergroup relations (Süssenbach & Gollwitzer, 2015), and might hence as well influence double standards in intergroup conflicts. We also focused on evaluations of immoral behaviors. Even though hypocrisy has been found for prosocial behaviors as well (Polman & Ruttan, 2012), we speculate that distrust predominantly affects moral standards for transgressions: there is a generally greater self-other dissociation in the immoral domain which is based on asymmetrical perceptions of intentions that distrust, in turn, should amplify (Gollwitzer et al., 2013; Klein & Epley, 2017).

Further, the present results were exclusively obtained with U.S. American samples on MTurk. This potential limitation notwithstanding, compared to typical student samples, MTurk samples are slightly more diverse (Buhrmester, Kwang, & Gosling, 2011) and more attentive (Hauser & Schwarz, 2016). To ensure sufficient attentiveness and hence data quality, we required a 95% approval rate for our participants (Peer, Vosgerau, & Acquisti, 2014), above and beyond attention and manipulation checks. Nevertheless, an investigation of the current hypotheses particularly in a population from a different cultural background could provide a valuable addition to the present set of studies.

Theoretical Contributions

The present findings extend the literature on moral hypocrisy, demonstrating how distrust can make moral cognition more flexible. Even though double moral standards may have highly detrimental consequences for interpersonal relationships and societies, only little research to date has addressed its antecedents. Notably, previous research has largely focused on the effects of approach-oriented variables on hypocrisy, such as social power and anger (Lammers et al., 2010; Polman & Ruttan, 2012). The present research is hence the first to investigate how a prevention-focused state of mind such as the experience of distrust affects moral standards for the self and others in differential ways.

In line with this perspective, distrust (vs. trust) overall appeared to have a stronger effect on judgments of others' transgressions, compared to judgments of one's own transgressions. As argued above, both lenient standards for the self and strict moral standards for others can serve to protect one's self-interest from exploitation (Gollwitzer et al., 2013). Indeed, previous research has attributed defective behavior and cheating not only to greed or a temptation by rewards, but also to the avoidance of exploitation or even regret (DeCremer, 1999; Effron et al., 2011)—which may then be rationalized in order to maintain a moral self-image. However, the present studies suggest that distrusting individuals may focus more strongly on preventing others' transgressions than achieving a competitive advantage. Furthermore, in a prevention orientation, individuals should focus more strongly on the potentially negative consequences of their own transgressions, such as punishment or reputational hazard, counteracting leniency for the self. Consequently, the processes underlying distrustful individuals' hypocrisy may differ from the reward motivation and entitlement that arguably drive the effect of power on hypocrisy (Lammers et al., 2010).

This perspective is in line with the moderating role of interindividual differences, which supports the notion that self-protective concerns, rather than increased entitlement or a general concern for justice and reciprocity, underlie the effect of distrust on moral standards.

Specifically, we found that the dispositional need to avoid exploitation amplified the effect of a target's untrustworthiness on hypocrisy. Earlier research has contrasted victim sensitivity's positive association with self-reported transgressions to its positive association with judgments of moral wrongness of the same behavior (Gollwitzer et al., 2005, Study 3). The present research extends these findings, specifying distrust as a situational condition under which victim-sensitive individuals may tend towards a self-serving morality. In addition, comparing moral judgments of identical behaviors if committed by the self versus others provides novel and more direct evidence for victim-sensitive individuals' hypocritical moral standards. As the latter may serve as justifications for subsequent behavior, the present work complements previous findings on the role of rationalization in the effects of victim sensitivity in close relationships (Gerlach et al., 2012): victim-sensitive individuals' lower forgiveness of close others' transgressions is associated both with mistrustful interpretations and with legitimizing cognitions regarding their own self-protective and unrelenting reactions. The present approach also extends research focusing on participants' own behaviors (e.g., Rothmund et al., 2011). It thus integrates previous research on the dispositional motivation to avoid exploitation with the effects of trust and distrust experiences, thereby providing additional empirical evidence for the SeMI Model (Gollwitzer et al., 2013).

Furthermore, the present findings tie in with accounts from the moral licensing domain: a trusted target may have accumulated enough moral "credit" to have some leeway for immoral behavior (Miller & Effron, 2010). Similarly, following a moral credentials account, a distrusted target may have less "moral credentials," leading to less favorable interpretations of ambiguous behaviors (Miller & Effron, 2010). While these arguments are less applicable to the carry-over effect in Study 3, expecting generally malevolent behavior from other persons may induce perceived license to transgress in distrusting individuals.

Moreover, the risk of being the "sucker" should threaten people's sense of competence (Vohs et al., 2007), which, in turn, can motivate self-enhancement in the moral domain (the "sucker-to-saint effect", Jordan & Monin, 2008). Plausibly, in order to deal with this threat to their agency, people may reframe their own transgressions as moral, or view others' as more distinct and immoral.

Interestingly, in the trust conditions of Study 4, participants were stricter with themselves than with another person, a phenomenon that has been coined *hypercrisy* (Lammers et al., 2010). Under trust, a reduced subjective need to protect one's self-interest and an enhanced concern for another's interests may promote leniency towards that target. Likewise, reduced entitlement and enhanced reputational concerns, or the risk of harming the relationship may attenuate the tendency to rationalize own transgressions. Hence, such reversed double standards may not just be a methodological artefact. Rather, the antecedents of hypercrisy, including relationship characteristics such as trust, deserve further investigation. Overall, however, the tendency for hypocrisy under distrust appeared to be stronger in the present data—potentially due to a general tendency for self-serving moral judgments (see Lammers et al., 2010, for a similar argument), and our focus on non-personal relationships.

More generally, the present research investigates the social-cognitive consequences of distrust in an inherently social domain, extending them to self-evaluations, and thereby contributing to existing work that largely focused on the cognitive characteristics of a distrust mindset or judgments about others (Kleiman et al., 2015; Posten & Mussweiler, 2013). Our findings also tie in with previous studies showing that subtle cues of distrust can increase cognitive flexibility, suggesting that distrust may as well affect flexibility of moral cognition. Indeed, flexible cognition may contribute to the present findings (Gino & Ariely, 2012); plausibly, distrustful individuals are better able to spontaneously entertain multiple,

alternative interpretations of moral situations and rules which helps them to flexibly focus on those aspects that support their preferred moral judgment (see Schul, Burnstein, & Bardi, 1996). In fact, recent research indicates that distrust can increase both utilitarian and deontological tendencies in moral dilemma judgments—suggesting more flexible and less one-sided moral cognition under distrust (Citation blinded, 2018). Lastly, experimental research on the social-motivational consequences of distrust is rare thus far. The current research documents the motivation to avoid exploitation as an important motivational moderator of the effects of distrust and thus complements recent findings indicating that distrust increases the motivation for control and predictability (Citation blinded, 2018). Conceivably, attempts to gain control and predictability over a situation may affect moral standards for the self and others (see also Lammers et al., 2010).

Importantly, hypocritical moral standards under distrust can affect behavior, promoting unscrupulous selfishness and harsh reactions to other's transgressions, for instance direct punishment or gossip that undermines the target's reputation. The present research hence provides insight into the psychological processes underlying previous findings that distrust leads to uncooperative behavior in social dilemmas such as public goods dilemmas (DeCremer, Snyder, & Dewitte, 2001; McCabe, Rigdon, & Smith, 2003). Furthermore, complementing an investigation of incidental distrust with distrust integral to a specific target and relationship, it elucidates how distrust may corrode interpersonal and intergroup relationships: once the same transgressions are not measured on the same scale anymore, and distrust increases further, conflicts become more difficult to reconcile. Moral hypocrisy may hence be an integral component of downward spirals of distrust and uncooperative behavior (Crocker & Canevello, 2015; Ferrin, Bligh, & Kohles, 2008).

Conclusion

Hypocrites do not practice what they preach, but send false signals about their own morality. Consequently, they evoke moral outrage and spark distrust (Jordan et al., 2017; Monin & Merrit, 2012). It is not without a sense of irony, then, that people who are distrustful of others' intentions may assume double moral standards in order to protect themselves, evaluating their own immoral behaviors more leniently than those of other persons. In this way, the present findings suggest that distrust can promote a dynamic that corrodes interpersonal relationships—with distrust provoking further hypocrisy, fueling deeper distrust among interaction partners.

References

Aquino, K., & Reed, A. I. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology, 72*, 1335–1348. http://doi.org/10.1037/0022-3514.72.6.1335

Baumert, A., & Schmitt, M. (2016). Justice sensitivity. In *Handbook of Social Justice Theory and Research* (pp. 161–180). New York, NY: Springer. http://doi.org/10.1007/978-1-4939-3216-0_9

Bocian, K., & Wojciszke, B. (2014). Unawareness of self-interest bias in moral judgments of others' behavior. *Polish Psychological Bulletin, 45*, 411-417.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5.

Burgmer, P. & Englich, B. (2013). Bullseye! How power improves motor performance. *Social Psychological and Personality Science, 4*, 224–232.

Cameron, C. D., & Payne, B. K. (2012). The cost of callousness: Regulating compassion influences the moral self-concept. *Psychological Science*, *23*, 225–229. http://doi.org/10.1177/0956797611430334

Carpenter, J., & Dolifka, D. (2017). Exploitation aversion: When financial incentives fail to motivate agents. *Journal of Economic Psychology*, 61, 213–224. http://doi.org/10.1016/j.joep.2017.04.006

Citation blinded. (2018). *Distrusting your moral compass: The impact of distrust mindsets on moral dilemma processing and judgments.* Manuscript in revision.

Citation blinded. (2018). *Malicious minds everywhere: Distrust, predictability motivation, and anthropomorphism.* Manuscript in preparation.

Crocker, J., & Canevello, A. (2015). Relationships and the self: Egosystem and ecosystem. In M. Mikulincer, P. R. Shaver, J. A. Simpson, & J. F. Dovidio (Eds.), *APA handbook of personality and social psychology, Volume 3: Interpersonal relations*. (pp. 93–116). Washington, DC: American Psychological Association. http://doi.org/10.1037/14344-004

De Cremer, D. (1999). Trust and fear of exploitation in a public goods dilemma. *Current Psychology*, *18*(2), 153–163. http://doi.org/10.1007/s12144-999-1024-0

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. http://doi.org/10.1002/ejsp.34

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 50, pp. 307–338). Philadelphia, PA: Elsevier. http://doi.org/10.1016/S0079-7421(08)00410-6

Effron, D. A., & Miller, D. T. (2011). Reducing exposure to trust-related risks to avoid self-blame. *Personality and Social Psychology Bulletin, 37*(2), 181–192. http://doi.org/10.1177/0146167210393532

Epley, N., & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research, 17*(2), 171–187. http://doi.org/10.1023/B:SORE.0000027408.72713.45

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*, 175-191

Ferrin, D. L., Bligh, M. C., & Kohles, J. C. (2008). It takes two to tango: An interdependence analysis of the spiraling of perceived trustworthiness and cooperation in interpersonal

and intergroup relationships. *Organizational Behavior and Human Decision Processes*, *107*, 161–178. http://doi.org/10.1016/j.obhdp.2008.02.012

Friesen, J., & Sinclair, L. (2011). Distrust and simultaneous activation of multiple categories. *Social Psychological and Personality Science*, *2*(1), 112–118. http://doi.org/10.1177/1948550610382666

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, *17*(12), 1068–1074. http://doi.org/10.1111/j.1467-9280.2006.01824.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. http://doi.org/10.1111/j.1467-6494.2012.00762.x

Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology*, *102*(3), 445–59. http://doi.org/10.1037/a0026406

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization*, *93*, 285–292. https://doi.org/10.1016/j.jebo.2013.04.005

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. http://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. http://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. http://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. http://doi.org/10.3389/fpsyg.2015.00439

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–68. https://doi.org/10.1037/a0034726

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366–385. http://dx.doi.org/10.1037/a0021847

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834. http://dx.doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences.* (pp. 852–870). Oxford, UK: Oxford University Press.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407. http://doi.org/10.3758/s13428-015-0578-z

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin*, *19*(5), 501–512. http://doi.org/10.1177/0146167293195003

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*(6202), 1340–1343. http://doi.org/10.1126/science.1251560

Jordan, A. H., & Monin, B. (2008). From sucker to saint: Moralization in response to self-threat. *Psychological Science*, *19*(8), 809–815. https://doi.org/10.1111/j.1467-9280.2008.02161.x

Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science, 28*(3), 356–368. http://doi.org/10.1177/0956797616685771

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science, 323*, 276-278.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*, 265–284. http://doi.org/10.1037/0033-295X.110.2.265

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, *13*(5), 505–521. http://doi.org/10.1080/026999399379168

Klein, N., & Epley, N. (2017). Less evil than you: Bounded self-righteousness in character inferences, emotional reactions, and behavioral extremes. *Personality and Social Psychology Bulletin*, 3(8) 1202–1212. http://doi.org/10.1177/0146167217711918

Kleiman, T., Sher, N., Elster, A., & Mayo, R. (2015). Accessibility is a matter of trust: Dispositional and contextual distrust blocks accessibility effects. *Cognition*, *142*, 333–344. http://doi.org/10.1016/j.cognition.2015.06.001

Lakens, D. (2014, June 7) *Calculating confidence intervals for Cohen's d and eta-squared using SPSS, R, and Stata.* [Blog post]. Retrieved from http://daniellakens.blogspot.de/2014/06/calculating-confidence-intervals-for.html

Lammers, J. (2012). Abstraction increases hypocrisy. *Journal of Experimental Social Psychology*, *48*(2), 475–480. https://doi.org/10.1016/j.jesp.2011.07.006

Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, *21*(5), 737–744. http://doi.org/10.1177/0956797610368810

Laurent, S. M., Clark, B. A. M., Walker, S., & Wiseman, K. D. (2014). Punishing hypocrisy: The roles of hypocrisy and moral emotions in deciding culpability and punishment of criminal and civil moral transgressors. *Cognition and Emotion*, *28*(1), 59–83. http://doi.org/10.1080/02699931.2013.801339

Leidner, B., & Castano, E. (2012). Morality shifting in the context of intergroup violence. *European Journal of Social Psychology*, *42*(1), 82–91. http://doi.org/10.1002/ejsp.846

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, *23*(3), 438–458. http://doi.org/10.5465/AMR.1998.926620

Mayer, J., & Mussweiler, T. (2011). Suspicious spirits, flexible minds: When distrust enhances creativity. *Journal of Personality and Social Psychology*, *101*, 1262–1277. http://doi.org/10.1037/a0024407

Mayo, R. (2015). Cognition is a matter of trust: Distrust tunes cognitive processes. *European Review of Social Psychology*, *26*(1), 283–327. http://doi.org/10.1080/10463283.2015.1117249

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267–275. http://doi.org/10.1016/S0167-2681(03)00003-9

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In Darcia Narvaez & Daniel K. Lapsley (Eds.), *Personality, Identity, and Character: Explorations in Moral Psychology* (pp. 341--354). New York: Cambridge.

Monin, B., & Merritt, A. (2012). Moral hypocrisy, moral inconsistency, and the struggle for moral integrity. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil.* (pp. 167–184). Washington, DC: American Psychological Association.

Peer, E., Vosgerau, J. & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*, 1023–1031 https://doi.org/10.3758/s13428-013-0434-y

Polman, E., & Ruttan, R. L. (2012). Effects of anger, guilt, and envy on moral hypocrisy. *Personality and Social Psychology Bulletin*, *38*(1), 129–139. http://doi.org/10.1177/0146167211422365

Posten, A.-C., & Mussweiler, T. (2013). When distrust frees your mind: The stereotype-reducing effects of distrust. *Journal of Personality and Social Psychology*, *105*, 567–584. http://doi.org/10.1037/a0033170

Rom, S. C., Weiss, A., & Conway, P. (2017). Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *Journal of Experimental Social Psychology*, *69*, 44–58. http://doi.org/10.1016/j.jesp.2016.09.007

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*, 107–119. https://doi.org/10.1177/0146167210391103

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment*, *21*(3), 202–211. http://doi.org/10.1027/1015-5759.21.3.202

Schul, Y., Burnstein, E., & Bardi, A. (1996). Dealing with deceptions that are difficult to detect : Encoding and judgment as a function of preparing to receive invalid information. *Journal of Experimental Social Psychology*, *253*(32), 228–253. http://doi.org/10.1006/jesp.1996.0011

Schul, Y., Mayo, R., & Burnstein, E. (2004). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology*, *86*, 668–679. http://doi.org/10.1037/0022-3514.86.5.668

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science, 24*(2), 125–130. https://doi.org/10.1177/0963721414553264

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, *37*(3), 330–349. http://doi.org/10.1177/0146167211398138

Steiger, J. H. (2004). Beyond the F Test: Effect Size Confidence Intervals and Tests of Close Fit in the Analysis of Variance and Contrast Analysis. *Psychological Methods, 9*(2), 164–182. http://doi.org/10.1037/1082-989X.9.2.164

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–71. http://doi.org/10.1016/j.cognition.2013.12.005

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. http://doi.org/10.1177/1368430214556700

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81. https://doi.org/10.1177/1745691614556679

Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, *44*(5), 1334–1338. https://doi.org/10.1016/j.jesp.2008.03.010

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

Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2017). Death before dishonor. *Social Psychological and Personality Science*, 194855061772027. http://doi.org/10.1177/1948550617720271

Yamagishi, T., Matsumoto, Y., Kiyonari, T., Takagishi, H., Li, Y., Kanai, R., & Sakagami, M. (2017). Response time in economic games reflects different types of decision conflict for prosocial and proself individuals. *Proceedings of the National Academy of Sciences*, *114*(24), 6394–6399. https://doi.org/10.1073/pnas.1608877114

Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly, 51*(3), 265–271. http://dx.doi.org/10.2307/2786924