Conway, P., Weiss, A., Burgmer, P., & Mussweiler, T. (2018). Distrusting your moral compass: The impact of distrust mindsets on moral dilemma processing and judgments. *Social Cognition*, *36*, 345-380. doi:10.1521/soco.2018.36.3.345

accepted version before copy-editing

Distrusting Your Moral Compass:

The Impact of Distrust Mindsets on Moral Dilemma Processing and Judgments

Paul Conway,[ab] Alexa Weiss,[b] Pascal Burgmer, [b] and Thomas Mussweiler [c]

[a] *Florida State University, USA*

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

[c] *London Business School, United Kingdom*

Word count: 9503

Research Article

Author Note: Correspondence concerning this article should be sent to Paul Conway, Florida State University, Department of Psychology, 1107 W. Call St., Tallahassee, Florida, USA, 32306-4301. Phone: (850) 644-4079. Email: conway@psy.fsu.edu

Abstract

A growing literature suggests that generalized distrust mindsets encourage carefully considering alternatives—yet it remains unclear whether this pertains to moral decision-making. We propose that distrust simultaneously increases opposing moral response inclinations when moral decisions pit two moral responses against one another, such as classic moral dilemmas where causing harm maximizes outcomes. Such a pattern may be invisible to conventional analytic techniques that treat dilemma response inclinations as diametric opposites. Therefore, we employed process dissociation to independently assess response inclinations underlying moral dilemma responses. Three studies demonstrated that activating generalized distrust (vs. trust and control) mindsets increased both harm avoidance and outcome-maximization response tendencies. These effects canceled out for conventional relative dilemma judgments. Moreover, perceptions of feeling torn between available response options mediated the impact of distrust on both response inclinations. These findings clarify how distrust impacts decision-making processes in the moral domain.

Keywords: trust, distrust, moral judgment, dilemmas, process dissociation

Distrusting Your Moral Compass:

The Impact of Distrust Mindsets on Moral Dilemma Processing and Judgments

"Mistrust first impulses; they are nearly always good."

—Charles Maurice de Talleyrand, Diplomat (1754-1838)

"The people I distrust most are those who want to improve our lives but have only one course of action."

—Frank Herbert, Novelist (1920-1986)

A host of research suggests that distrust can operate as a generalized mindset that shakes people out of default information processing by increasing the salience of relevant alternative possibilities (e.g., Schul, Mayo, & Burnstein, 2004). Such work has intriguing, yet unexplored, implications for moral psychology, and moral decision-making in particular: when considering classic moral dilemmas where causing harm maximizes outcomes, such as harming research animals to cure AIDS, distrust may increase people's desire to both response options. Yet, when deliberating between clearly moral response options (e.g., refusing to harm research animals) versus more self-interested options (e.g., harming animals to improve beauty products), distrust might reduce people's tendency to make moral decisions by highlighting the tempting non-moral alternative. Hence, inducing generalized distrust might either increase or reduce moral response inclinations depending on which response alternatives people consider. We present three studies demonstrating this phenomenon, exploring how distrust alters moral decision-making processes.

Trust/Distrust Mindsets

Although trust and distrust sometimes pertain to specific targets, they may also operate as generalized mindsets that influence whatever decisions are currently under consideration (Schul, Mayo, & Burnstein, 2004). A host of research indicates that trust mindsets signal that all is well with the world, so careful processing is unnecessary; one may uncritically rely on routine processing to get by. Trust indicates that situations, people, and appearances can be taken at face value, as their hidden features match their surface properties. Conversely, distrust mindsets signal that something is amiss in the environment and it therefore warrants close scrutiny; one should carefully consider alternatives to one's initial conclusions (e.g., Kleiman et al., 2015; Mayer & Mussweiler, 2011; Mayo, Alfasi, & Schwarz, 2014; Schul, Mayo, & Burnstein, 2004; Schul, Mayo, & Burnstein, 2008).

Accordingly, distrust alters information encoding and processing to help resolve suspicious situations and avoid being misled: generalized distrust mindsets are characterized by questioning ones' default position and engaging in non-routine information processing (Schul et al., 2008; Mayer & Mussweiler, 2011; Posten & Mussweiler, 2013). Specifically, people in a distrust mindset tend to apply multiple interpretive frames to a given situation or set of information—that is, they consider events from multiple perspectives—and to refrain from inferring dispositions from behaviors (Fein, 1996). For example, under distrust, people encode messages as both true and false (Schul, Burnstein, & Bardi, 1996), apply multiple information categories (Friesen & Sinclair, 2011), remain vigilant for unusual contingencies (Schul et al., 2008), apply more flexible and creative problem-solving strategies (Mayer & Mussweiler, 2011), reduce reliance on stereotypes in favor of individuating information (Posten & Mussweiler, 2013), and engage in more disconfirmatory hypothesis testing (Mayo, Alfasi, & Schwarz, 2014). Moreover, stimuli encountered under distrust activate incongruent

associations—for example, *black* activates *white* and *hollow* activates *full* (e.g., Schul et al., 2004). Furthermore, distrust has been shown to block accessibility effects in several domains, such as person perception (Kleiman et al., 2015). Generalized trust and distrust mindsets hence alter processing strategies applied to information unrelated to the initial distrust-eliciting stimulus (e.g., Schul et al., 2004). In sum, whereas generalized trust mindsets entail routine information processing and uncritically accepting default positions, distrust mindsets promote questioning one's default and considering alternative responses and interpretations. Therefore, when faced with a difficult choice as in moral dilemmas, people in a distrust mindset should feel more torn between whatever response options are available. Considering one possible response should activate the opposing response as well (Kleiman et al., 2015).[1]

Distrust mindsets can be activated by providing subtle cues hinting that people may have ulterior motives or that objects may be unreliable. Researchers have employed a variety of methods to do so, including activating semantic concepts related to distrust, confronting participants with untrustworthy social targets, or asking participants to recall experiences of distrust (e.g., Kleiman et al, 2015; Mayer & Mussweiler, 2011). Such manipulations are typically contrasted against manipulations designed to activate generalized trust mindsets, which hint that objects and situations are reliable or that people may have honest and forthright motives, or neutral conditions that involve processing ordinary information. When people encounter morally-relevant decisions, we hypothesized that inducing distrust may jostle people out of making their default decision by inducing them to fully consider the alternative response, thereby ratcheting up their desire to select both response options (i.e., decisional ambivalence). Whereas such an effect may be invisible to classic dilemma analytic techniques, it may nonetheless be detectable via process dissociation.

Moral Dilemmas

Philosophers originally developed moral dilemmas to contrast two responses that correspond to philosophical positions (Foot, 1967). In classic dilemmas, causing harm maximizes overall outcomes, such as the dilemma where testing harmful chemicals on animals helps to cure AIDS. *Deontological* ethical positions entail rejecting harm, because they define morality by the intrinsic nature of an action: harm violates individuals and is therefore immoral regardless of what good it might achieve (Kant, 1875/1959). Conversely, *utilitarian* ethical positions entail accepting harm to maximize outcomes, because they define morality by the consequences of actions: actions that maximize net outcomes (utility) are moral even if they involve causing harm (Mill, 1861/1998). Accordingly, some theorists have referred to harm rejection as the 'characteristically deontological' judgment, and harm acceptance as the 'characteristically utilitarian' judgment (e.g., Greene et al., 2001). Note that this does not mean that judgments were *caused by* the philosophies in question; rather they are *consistent with* those philosophies. Hence, referring to such judgments as 'utilitarian' or 'deontological' can be problematic (see Kahane, 2015); nonetheless we retain this terminology in order to maximize consistency with past work.

The most prominent model of moral dilemma responses is the dual-process model of moral judgment (Greene et al., 2001; c.f., Mikhail, 2007), which postulates that moral dilemma responses stem from the competition between two psychological processes: an affect-laden aversion to causing harm, which motivates harm rejection (despite the lives lost), and a cognitive evaluation of outcomes, which motivates harm acceptance (in order to save lives). Although the original conceptualization suggested that affective responses precede cognitive ones in processing (Greene et al., 2001), reaction time data supporting this

conception has been withdrawn (Greene et al., 2004) and recent findings suggest that each response tendency requires a similar amount of time to process (e.g., Baron & Gürçay, 2016). Therefore, the dual-process model is better conceptualized as a competition between relatively affect-laden tendencies to reject harm and relatively cognitive tendencies to accept harm to maximize outcomes that require similar processing time. Although experimental manipulations certainly influence responses (e.g., Amit & Greene, 2012), people typically prefer one or the other response by default (Helzer et al., 2016), and stable individual differences reliably predict dilemma judgments (e.g., Bartels, 2008). Thus, not all people may experience strong conflict when encountering dilemmas; many may simply select their default dilemma choice unless they have reason to carefully consider the alternative.

Typically, moral dilemma researchers confine investigation only to classic dilemmas where causing harm maximizes outcomes. Although such studies generally support the dual-process model, they suffer from an important methodological limitation: they treat deontological and utilitarian judgments as perfect opposites. Participants must either endorse or reject causing outcome-maximizing harm, such that higher scores reflect relatively more endorsement of utilitarian responses and lower scores reflect relatively more endorsement of deontological responses. Thus, the measurement of deontology and utilitarianism is not independent. Yet, the model posits two independent processes that jointly contribute to conventional relative judgments (Greene et al., 2001). Using classic methods, any factor that simultaneously increases both response tendencies would cancel out, much like adding weight evenly to both sides of a scale retains its balance even as the absolute amount of weight changes. Accordingly, two people may select the same conventional dilemma response for very different reasons. One might feel strong inclinations to select both

responses, and consequently feel torn between them, before ultimately selecting only one. Another might feel weak inclinations to select either option, and consequently not feel torn between them, yet arrive at the same judgment. Conventional classic dilemmas can hence not distinguish between these possibilities due to the non-independent measurement of deontological and utilitarian responses.

Other predictions are also possible. First, if generalized distrust operates similar to personal distrust, distrust manipulations may undermine confidence in evidence (e.g., Darke & Ritchie, 2007; Schul & Peri, 2015). In this case, evidence pertains to the dilemma information provided by researchers. Distrust in evidence may result in participants doubting that actions will cause harm, or doubting that actions will succeed in causing beneficial actions. Either way, sowing such distrust ought to reduce one or both response tendencies: if actions will not cause harm, there is no need to avoid them, if actions will not improve the world, there is no need to perform them. Second, insofar as the distrust items are more negative than the trust items, they may induce negative emotion or a general negativity bias. Past work suggests negative emotions selectively increase deontological responding (e.g., Valdesolo & DeSteno, 2006), as deontological responses are tied more directly to emotional concern for others (Conway & Gawronski, 2013). Alternatively, general negativity may lead to a general increase in harm acceptance, which would translate into reductions on both parameters. Hence, this prediction would suggest that distrust mindsets should either selectively increase deontological inclinations, or should decrease both deontological and utilitarian inclinations. The current work has potential to distinguish between these possibilities and our focal prediction that distrust will increase both deontological and utilitarian response tendencies.

Process Dissociation

To overcome the non-independent measurement of deontology and utilitarianism endemic to conventional dilemma measures, Conway and Gawronski (2013) adapted a technique called process dissociation (Jacoby, 1991) to independently assess deontological and utilitarian response inclinations underpinning conventional dilemma judgments. Process dissociation is a widely-applied content-agnostic procedure for distinguishing the independent contributions of multiple processes theorized to jointly contribute to dichotomous judgments (Payne & Bishara, 2009). The key insight was to measure responses on both trials that pit the two underlying processes against one another (incongruent trials) as well as trials where the two underlying processes lead to the same response (congruent trials).

Conway and Gawronski (2013) developed a new battery composed of both incongruent and congruent moral dilemmas. *Incongruent* dilemmas corresponded to classic, conventional, high-conflict dilemmas because causing harm maximizes overall outcomes (see Koenigs et al., 2007)—hence, deontological and utilitarian responses are incongruent. *Congruent* dilemmas are worded identically to their incongruent counterparts, except that the outcome of harm is reduced, such that causing harm no longer maximizes overall outcomes. For such dilemmas, rejecting harm is consistent with both deontology and utilitarianism—that is, these inclinations are congruent. However, people may accept causing harm on congruent dilemmas for other (non-moral) reasons, such as self-interest, vengeance, or sadism. By applying response patterns to both kinds of dilemmas to a decision tree, the unique component of each process can be derived mathematically (see Figure 1), and researchers can derive two parameters: one tapping ‘deontological’ inclinations to

consistently reject harm, and the other taping 'utilitarian' inclinations to consistently maximize outcomes.

A growing body of work supports the utility of the process dissociation approach. Empirically, these parameters are largely uncorrelated, yet both correlate sensibly with conventional relative judgments: people higher on the utilitarian parameter tend to make more utilitarian versus deontological conventional judgments, whereas people higher on the deontological parameter tend to make more deontological than utilitarian conventional judgments. Friesdorf, Conway, and Gawronski (2015) confirmed this pattern meta-analytically across 40 datasets. In line with the dual-process model, the deontological parameter appears to tap primarily processes related to affective reactions to harm, whereas the utilitarian parameter appears to tap primarily cognitive processes related to cost-benefit calculations, as manipulations can impact them individually and they have different impacts on related variables (Conway & Gawronski, 2013; Park, Kappes, Van Bavel, & Rho, 2016; Lee & Gino, 2015).

However, for present purposes the distinction between affective and cognitive processing is less important than the fact that the parameters vary independently. As a result, some constructs may increase both parameters simultaneously. For example, Conway and Gawronski (2013) found that moral identity—the degree to which morality is central to the self-concept (Aquino & Reed, 2002)—predicted both the deontological and utilitarian parameters. These positive relations cancelled out for conventional dilemma judgments when deontological inclinations were pitted directly against utilitarian ones. We anticipated obtaining a similar effect for distrust: by jostling people out of their default judgment preference and inducing them to more carefully consider the alternative response, and then in

turn reconsider their initial response, distrust should increase both parameters. However, this simultaneous increase should cancel out for conventional judgments. We also expected an opposite effect for trust mindset manipulations: we anticipated that inducing a generalized trust mindset might lull participants into increased security in their initial preference, leading to a reduction in both parameters as participants need not scrutinize either response carefully in order to complete the dilemma task. They could simply select one without considering the other (with some people uncritically selecting one answer and other people uncritically selecting the other answer, leading to a lower net level of either answer compared to the distrust condition where people ratchet up their desire to select both by vacillating between them. Moreover, we expected a neutral condition to emulate the trust condition, based on work suggesting that trust mindsets operate as the default (e.g., Mayo, 2015; Schul et al., 2004). Calculating the deontology and utilitarian PD parameters requires examining harm acceptance and rejection judgments for both congruent and incongruent dilemmas. As harmful action maximizes outcomes in incongruent but not congruent dilemmas, accepting harm on incongruent dilemmas but rejecting harm on congruent dilemmas upholds utilitarianism. Conversely, deontology entails avoiding causing harm, so rejecting harm always upholds deontology. We illustrate this point in Figure 1: the top path illustrates the case where utilitarianism drives dilemma judgments—accepting harm for incongruent and harm for congruent dilemmas. The second path illustrates the case where deontology drives dilemma judgments—thus always rejecting harm. The bottom path presents the case where neither utilitarianism nor deontology drives dilemma responses, such that causing harm is acceptable because people have neither utilitarian nor deontological reasons to avoid it.

By reviewing the two rightmost columns in Figure 1, researchers can derive which cases led participants to reject or accept harm for congruent and incongruent dilemmas. For congruent dilemmas, harm rejection occurs either when utilitarianism drives responses: U, or when deontology drives responses: $(1 - U) \times D$. Conversely, harm acceptance for congruent dilemmas occurs when neither utilitarianism nor deontology drives responses: $(1 - U) \times (1 - D)$. For incongruent dilemmas, harm rejection occurs when deontology drives responses: $(1 - U) \times D$. Conversely, harm acceptance occurs when either utilitarianism drives responses: U, or when neither utilitarianism nor deontology drives responses: $(1 - U) \times (1 - D)$.

By mathematically representing the probability of each case and plugging in each participants' decisions, researchers can combine and rearrange equations to algebraically solve for the two unknown parameters (D and U) for each participant. For example, the probability of rejecting harm on congruent dilemmas involves either the case where utilitarianism drives responses, or deontology drives responses:

$$\text{Eq. (1) } p(\text{unacceptable} \mid \text{congruent}) = U + [(1 - U) \times D]$$

Conversely, the probability of accepting harm for congruent dilemmas involves the case when neither utilitarianism nor deontology drives responses:

$$\text{Eq. (2) } p(\text{acceptable} \mid \text{congruent}) = (1 - U) \times (1 - D)$$

For incongruent dilemmas, the probability of rejecting harm occurs when deontology drives responses:

$$\text{Eq. (3) } p(\text{unacceptable} \mid \text{incongruent}) = (1 - U) \times D$$

Conversely, the probability of rejecting harm occurs when either utilitarianism drives responses, or neither deontology nor utilitarianism drives responses:

$$\text{Eq. (4) } p(\text{acceptable} \mid \text{incongruent}) = U + [(1 - U) \times (1 - D)]$$

By representing all probabilities algebraically for congruent and incongruent dilemmas, researchers can plug in each participants' pattern of actual responses and algebraically combine these equations to solve for two parameters estimating deontological (*D*) and utilitarian (*U*) inclinations underpinning their responses. Specifically, by combining Equations 3 and 1, researchers can solve for *U:*

Eq. (5) U = *p*(unacceptable | congruent) - *p*(unacceptable | incongruent)

Once *U* has been obtained, researchers can plug in this value to Equation 3 to solve for *D* thus:

Eq. (6) D = *p*(unacceptable | incongruent) / (1 – U)

Together, these formulas enable researchers to derive two parameters that independently estimate the strength of deontological and utilitarian inclinations underlying conventional relative dilemma judgments (see Conway & Gawronski, 2013, for details). [2]

Hypotheses and Overview of Current Work

In light of existing evidence demonstrating that generalized distrust increases consideration of alternatives and non-routine processing, we anticipated that activating distrust (versus trust) mindsets will jostle people out of their default dilemma response strategy by inducing them to carefully consider the alternative response. Due to the specific information-processing strategies induced by distrust, distrustful participants should be inclined to select both responses—that is, demonstrate increases in both the deontological and utilitarian process dissociation parameters underpinning conventional dilemma judgments. These simultaneous increases should cancel out for conventional relative dilemma judgments that pit deontology directly against utilitarianism. Moreover, considering and contrasting both response options should increase experienced conflict during dilemma

decision-making. Thus, we expect that perceptions of feeling torn (i.e., ambivalent) between available dilemma response options would mediate the effect of distrust on dilemma processing. Conversely, we anticipated that inducing generalized trust would lull participants into heuristically and uncritically relying on whichever initial impulse they typically prefer, rather than ambivalently considering both responses. We also anticipated that a neutral condition would emulate the trust condition. We report three studies investigating these predictions, and provide a meta-analysis including three additional file-drawer studies (see Supplementary Material). We report all exclusions, independent and dependent variables, sample size and data-stopping decisions for all studies. Data and analyses for all studies are available at osf.io/25mhq.

Study 1

In Study 1 we activated either trust or distrust mindsets via a scrambled-sentences task adapted from previous work (e.g., Mayer & Mussweiler, 2011). Participants in the trust-mindset condition unscrambled sentences describing trustworthiness, honesty, reliability, etc., whereas those in the distrust-mindset condition unscrambled sentences regarding untrustworthiness, dishonesty, unreliability, etc. In between unscrambling sentences, participants completed the moral dilemma battery from Conway and Gawronski (2013). This battery allows for assessing conventional deontological versus utilitarian dilemma judgments, consistent with standard practice (e.g., Greene et al, 2001). However, this battery also allows for calculating independent estimates of the deontological and utilitarian inclinations underpinning participants' dilemma judgments. We examined how trust versus distrust mindsets impacted both conventional judgments and each parameter. We predicted that

activating generalized distrust (relative to trust) would increase both parameters, but that conventional judgments would remain unaffected.

Method

Participants. We recruited 89 American participants (36 female, 53 male, M_{age} = 33.09, *SD* = 10.83) for payment via Amazon's Mechanical Turk (Amazon, 2015). A recent meta-analysis on the process dissociation of moral dilemmas found no differences between Mechanical Turk and laboratory samples (Friesdorf et al., 2015; see also Buhrmester, Kwang, & Gosling, 2011). We aimed for approximately 50 participants per cell (see Simmons, Nelson, & Simonsohn, 2011), but ended up with slightly fewer. A post-hoc power analysis using GPower indicated this sample provided ~87% power to detect the main effect of mindset across both parameters (Faul, Erdfelder, Lang, & Buchner, 2007). Data were collected in a single run. There were no exclusions.

Procedure. Participants completed all measures online. First, they read instructions pertaining to the scrambled-sentences task as well as the dilemma task, and were told the tasks were interspersed. Then participants unscrambled 10 sentences (5 experimental, 5 control) before encountering the first moral dilemma. Thereafter, participants unscrambled two sentences (1 experimental, 1 control, in a fixed randomized order) in between answering each of 20 moral dilemmas. This procedure allowed for repeatedly re-activating concepts related to trust/distrust, while the true purpose of the manipulation was masked by filler sentences.

Scrambled-sentences task. We adapted the scrambled-sentences task from past work on this topic (Mayer & Mussweiler, 2011; Posten & Mussweiler, 2013). For each scrambled sentence, participants were presented with five words in a random order (all capitalized) and

were asked to enter a valid four-word sentence into a text box (as one word was always irrelevant). For example, *COOKIES HAT SOME THEY BAKE* could be unscrambled as *They bake some cookies*. Participants in the trust-mindset condition unscrambled 24 sentences regarding trustworthiness, honesty, and reliability (e.g., *The car starts reliably; Her story was believable*). Participants in the distrust-mindset condition unscrambled 24 sentences regarding untrustworthiness, dishonesty, and unreliability (e.g., *He asked misleading questions; The path seems uncertain*). Scrambled sentences from all studies are available in Appendix A. We interspersed the priming sentences in each condition with 24 neutral sentences (e.g., *It is sunny today; They saw her walking*) to mask the goal of the manipulation. This mask appeared effective: only two participants identified a theme of trust or distrust in open-ended responses, and one additional person who identified the purpose of the study. Results remain similar when excluding them.[3]

Moral dilemma task. Next, participants completed 10 moral dilemmas, each with one congruent and one incongruent version, in the same fixed random order as in Conway and Gawronski (2013). *Incongruent* dilemmas correspond to *high-conflict* (Koenigs et al., 2007) moral dilemmas where causing harm leads to the best overall outcome. For example, in the incongruent crying baby dilemma, participants must decide whether it is acceptable to kill an infant to prevent its cries from attracting murderous soldiers who would kill all the townsfolk. Such dilemmas are said to pit deontological against utilitarian ethical considerations (Foot, 1967). *Congruent* dilemmas employ similar wording to incongruent dilemmas, except the outcome of causing harm is reduced such that causing harm no longer leads to the best overall outcome. For example, in the congruent crying baby dilemma, participants must decide whether it is acceptable to kill an infant to prevent its cries from

attracting soldiers who will merely cause the townsfolk to labor in a mine. As causing harm no longer leads to the best overall outcome, deontological and utilitarian ethical positions are no longer in conflict—harm rejection is consistent with *both* deontology and utilitarianism. Note, however, that people may still accept causing harm for other reasons (e.g., pragmatism, selfishness, vengeance, etc.).

For each dilemma, participants indicated whether causing harm is *appropriate* or *not appropriate* (see Greene et al., 2001). For example, is it appropriate to kill the baby to save the other townsfolk? Conventional dilemma analyses examine only responses to incongruent dilemmas, where the proportion of times participants indicate that causing harm is acceptable corresponds to conventional utilitarian versus deontological judgments, and the inverse proportion (where participants indicated that causing harm is unacceptable) corresponds to conventional deontological versus utilitarian judgments. As these figures are the inverse of one another, we present only the findings for judging harm as relatively acceptable (i.e., relatively more utilitarian versus deontological judgments). Process dissociation allows for the analysis of not only conventional relative moral dilemma judgments, but also two independent parameters reflecting the strength of deontological and utilitarian inclinations: the deontology and utilitarian parameters. We followed the six steps required to calculate the parameters as described by Conway and Gawronski (described above).

Results

Conventional dilemma analysis. First, we examined whether participants in the trust or distrust mindset conditions were more likely to accept causing harm on conventional, incongruent, high-conflict moral dilemmas where harm maximizes outcomes. As predicted, no significant differences emerged between the distrust (M = .60, SD = .19) and trust (M =

.62, *SD* = .21) conditions, $t(87) = -0.35$, $p = .725$, 95% CI_{diff} [-.10, .07], and the effect size was very small: Cohen's d = -.09. As null effects can be difficult to interpret, we computed the Scaled-Information Bayes Factor in favor of the null where scale $r = .707$ (Rouder, Speckman, Sun, Morey & Iverson, 2009). This Bayes factor was 3.22, suggesting that the data were about 3 times more likely to support the null than alternative hypothesis.

Process dissociation analysis. Next, we computed the process dissociation parameters. As depicted in Table 1, both parameters correlated sensibly with conventional relative dilemma judgments, but correlated only mildly with one another, consistent with recent meta-analytic findings (Friesdorf et al., 2015). As the parameters are on different scales, we always standardized them prior to analysis to eliminate the theoretically meaningless main effect of scale. However, analyses using unstandardized parameters obtain the same patterns.

We conducted a 2 (mindset: distrust versus trust) × 2 (parameter: utilitarian versus deontological) mixed-model ANOVA with the first factor between-subjects and the second factor within-subjects. The results of this analysis revealed no main effect of parameter, $F(1, 87) = .01$, $p = .919$, $\eta_p^2 < .001$, and no interaction, $F(1, 87) = .23$, $p = .634$, $\eta_p^2 = .003$ (see Figure 2). However, we obtained the predicted main effect of mindset: Participants in a distrust mindset (*M* = .27, *SD* = .87) scored higher on both parameters than those in a trust mindset (*M* = -.18, *SD* = .67), $F(1, 87) = 7.48$, $p = .008$, $\eta_p^2 = .079$. Post-hoc tests confirmed that the deontology parameter was significantly higher in the distrust (*M* = .31, *SD* = .98) than trust condition, (*M* = -.20, *SD* = .97), $F(1, 87) = 5.89$, $p = .017$, $\eta_p^2 = .063$, 95% CI_{diff} .09, .93], and the utilitarian parameter was marginally higher in the distrust (*M* = .23, *SD* = 1.15) than trust condition (*M* = -.15, *SD* = .87), $F(1, 87) = 3.19$, $p = .078$, $\eta_p^2 = .035$, 95%

CI_{diff} [-.04, .81]. These findings remained significant when gender was included as a covariate.[4]

Discussion

Although no difference between the trust and distrust mindsets emerged for classic dilemma judgments, process dissociation revealed that participants in the distrust condition scored higher on both parameters—indicating they experienced both stronger inclinations to avoid causing harm (deontological inclinations) and stronger inclinations to maximize outcomes (utilitarian inclinations) than did participants in a trust mindset. These findings suggest that activating distrust jostled participants out of their default dilemma response strategy by inducing them to more carefully consider the alternative response, and vice versa, thereby ratcheting up their desire to select both responses. These simultaneous increases cancelled out for conventional judgments, much like the simultaneous positive effects of moral identity on both parameters cancelled out for conventional judgments in Conway and Gawronski (2013).

Study 2

Although Study 1 offers initial evidence for our argument, conducted an exact replication to determine whether this effect is reliable. We also added a control group who read only neutral sentences. We anticipated replicating the difference obtained between distrust and trust conditions obtained in Study 1. Based on research suggesting that people operate in trust mindsets by default (e.g., Mayo, 2015; Schul et al., 2004), we also anticipated that a difference might emerge between the neutral and distrust condition, whereas no such difference might emerge between the neutral and trust condition. Hence, we anticipated that relative to distrust, people in the neutral and trust conditions would score lower on both

parameters, as they would be lulled into uncritically selecting a response rather than vacillating between responses with some people uncritically selecting one response, and others uncritically selecting the other response, leading to a lower net scores than under distrust mindsets where we theorized people vacillate.

Method

Participants. The final sample consisted of 164 (80 female, 84 male, M_{age} = 35.91, *SD* = 12.34) American Mechanical Turk workers for pay, after excluding 11 participants (distrust: 7, neutral: 0, trust: 4) who reported that their data were unreliable. Effects are similar, but approximately half the effect size, when analyzing the full sample. Based on the effects obtained in Study 1, we aimed for approximately 180 participants (60 per condition) during a single mturk run. A post-hoc power analysis indicated that this sample provided ~77% power to detect the main effect of mindset across both parameters.

Materials and procedure. Overall the procedure was the same as Study 1, except we added a control group. Participants in the trust and distrust conditions saw the same 25 trust or distrust sentences as in Study 1, interspersed with the same 25 filler neutral sentences. Participants in the neutral condition saw 25 additional neutral sentences (e.g., *They read the book*) in lieu of the trust/distrust sentences. All participants completed the same dilemma battery.[5]

Results

Conventional dilemma analysis. Again, we began by examining the proportion of people who accepted harm on incongruent dilemmas in each condition (i.e., who made relatively more utilitarian than deontological decisions). Replicating Study 1, an omnibus ANOVA revealed no significant difference between accepting harm in the distrust (*M* = .57,

SD = .19), neutral (*M* = .56, *SD* = .20), or trust (*M* = .60, *SD* = .23) mindset conditions for conventional relative moral dilemma judgments, $F(2, 161) = .75$, $p = .475$, $\eta_p^2 = .009$. The 95% CIs for the difference between both the distrust-neutral [-.08, .08] and neutral-trust conditions [-.12, .03] contained zero. The Scaled-Information Bayes Factor in favor of the null where scale $r = .707$ was 4.79, suggesting that the data were almost five times more likely to support the null than alternative hypothesis.

Process dissociation analysis. We again computed the process dissociation parameters, and entered them into a 3 (mindset: distrust versus neutral versus trust) × 2 (parameter: utilitarian versus deontological) mixed-model ANOVA with the first factor between-subjects and the second factor within-subjects. As before, the parameters correlated sensibly with conventional judgments and one another (see Table 1). This analysis revealed no main effect of parameter, $F(1, 161) = .01$, $p = .906$, $\eta_p^2 < .001$, and no interaction, $F(2, 161) = .62$, $p = .537$, $\eta_p^2 = .008$ (see Figure 3). However, we obtained a marginally significant main effect of mindset in the predicted direction, $F(2, 161) = 2.87$, $p = .059$, $\eta_p^2 = .034$. Pairwise comparisons indicated that the parameters were significantly higher in the distrust (*M* = .16, *SD* = .76) than trust condition, (*M* = -.17, *SD* = .79), $t = 2.30$, $p = .023$, $d = 0.43$. The neutral condition (*M* = .06, *SD* = .71) fell in between, and did not significantly differ from either the distrust, $t = -0.65$, $p = .513$, $d = -0.14$, or trust conditions, $t = 1.64$, $p = .103$, $d = 0.31$.

Considering each parameter separately, post-hoc tests confirmed that the deontology parameter was significantly higher in the distrust (*M* = .21, *SD* = .85) than trust condition (*M* = -.24, *SD* = 1.13), $t = 2.41$, $p = .017$, $d = 0.45$, 95% CI_{diff} [.08, .82], but there was no significant difference between the distrust and neutral (*M* = .01, *SD* = .91), $t = -0.56$, $p =$

.575, d = -0.23, 95% CI_{diff} [-.28, .50], or neutral and trust conditions, t = 1.84, p = .066, d = 0.24, 95% CI_{diff} [-.02, .71]. The utilitarian parameter itself did not vary significantly between the distrust (M = .11, SD = 1.03) and trust conditions, (M = -.10, SD = .98), t = 1.10, p = .271, d = 0.21, 95% CI_{diff} [-.17, .59], nor distrust versus neutral (M = .02, SD = 1.01), t = -0.44, p = .661, d = -0.09, 95% CI_{diff} [-.31, .48], nor neutral versus trust conditions, t = 0.65, p = .512, d = 0.12, 95% CI_{diff} [-.25, .49].

Discussion

Although Study 2 largely failed to reach statistical significance, it garnered a pattern of results broadly consistent with Study 1: activating distrust (versus trust) increased deontological inclinations, and there was a nonsignificant trend for utilitarian inclinations to move in the same direction. Scores in the neutral condition fell in between these extremes, albeit non-significantly so. Once again, conventional relative dilemma judgments did not vary significantly across mindset condition. However, we did not confirm the stronger hypotheses that distrust would increase both parameters relative to the control condition. Descriptively, the control condition fell between the trust and distrust conditions, albeit closer slightly to the trust condition, suggesting that perhaps trust mindsets approximate the default, and distrust mindsets depart from this default (Mayo, 2015; Schul et al., 2004). Due to insufficient power and non-significance of the result, such an interpretation remains speculative, however. These results were rather weak, but at the very least they do not contradict the findings of Study 1. These findings slightly increased our confidence in the claim that activating generalized distrust increases the desire to select both available dilemma responses. Still, we thought it wise to attempted to replicate this pattern again before drawing conclusions regarding how robust it is. Moreover, in Study 3, we assessed a potential

mechanism driving this effect: If participants in a distrust (versus trust) mindset experienced both stronger inclinations to reject harm and stronger inclinations to accept harm (to maximize outcomes), then they should feel more torn between response alternatives, which should mediate the effect of distrust on the parameters.

Study 3

Studies 1 and 2 demonstrated that people in a distrust mindset scored higher on both utilitarian and deontological dilemma response tendencies than people in a trust mindset. However, the mechanism behind this effect remains unclear. The distrust mindset literature indicates that distrust induces people to engage in non-routine processing and to consider alternatives (Schul et al., 2004, 2008; Posten & Mussweiler, 2013), as under distrust, considering one concept automatically activates alternatives (Kleiman et al., 2015). In addition, distrust enhances cognitive flexibility, possibly increasing creative consideration of implications of each dilemma judgment (Mayer & Mussweiler, 2011). Hence, we theorized that distrust leads participants to second-guess their default dilemma response, and carefully consider the alternative response—whereupon they may reconsider their default response. If this explanation is correct, then participants in a distrust (versus trust) mindset should feel more torn between the available response options, which should mediate the impact of the manipulation on the parameters.

Importantly, the *direction* of mediation should depend on which response alternatives people consider. Recall that for incongruent dilemmas, causing harm maximizes outcomes, so rejecting harm is consistent with deontology and accepting harm is consistent with utilitarianism. Accordingly, when distrust increases the desire to select both of the response options for incongruent dilemmas, people ought to score higher on both the deontological

and utilitarian parameters. In other words, feeling torn on incongruent dilemmas should *positively* mediate the effect of distrust on each parameter. Conversely, for congruent dilemmas, causing harm does not maximize outcomes. Therefore, rejecting harm is consistent with both deontology and utilitarianism, whereas accepting harm is consistent with neither (though still consistent with other, non-moral motives, such as self-interest, vengeance, and sadism). Hence, ambivalence on congruent dilemmas is effectively ambivalence between, on the one hand, an answer consistent with deontology *and* utilitarianism, versus, on the other hand, an answer consistent with neither (but confident with self-interest, vengeance, sadism, etc.). Therefore, ambivalence on congruent dilemmas reflects reduced deontological and utilitarian concerns versus other dilemma considerations. Accordingly, feeling torn on congruent dilemmas should *negatively* mediate the effect of distrust on each parameter, because such ambivalence reflects ambivalence between a moral (deontological/utilitarian) response and an amoral/immoral response.

Moreover, both of these mediation effects may occur simultaneously within the same analysis. Distrust may increase the desire to select both responses for incongruent dilemmas, increasing both moral responses, yet also increase desire to select both responses for congruent dilemmas, decreasing both moral responses by pitting them against an immoral alternative. So long as people feel more torn on incongruent than congruent dilemmas overall, the former effect will win out against the latter, resulting in a net increase in both parameters. We anticipated that people would indeed feel more torn between response options for incongruent than for congruent dilemmas, as past work has demonstrated that people find incongruent dilemmas more difficult to resolve than congruent ones (Conway & Gawronski, 2013; Friesdorf et al., 2015). We examined these possibilities in Study 3.[6]

Method

Participants. The sample consisted of 124 American Mechanical Turk workers for pay (65 female, 59 male, M_{age} = 36.20, *SD* = 10.77), after we excluded 10 participants (distrust: 4, trust: 6) who indicated either a) their data were unreliable using the same measure as Study 2 (3), or who reported completing the PD dilemma battery in previous unrelated research (7). Effects are actually stronger when analyzing the full sample. We again aimed for ~60 participants per cell in a single mturk run. A post-hoc power analysis indicated that this sample provided ~99% power to detect the main effect of mindset across both parameters.

Procedure. Again, we employed the scrambled-sentences task to induce trust or distrust mindsets and assessed responses to the Conway and Gawronski (2013) dilemma battery. However, this time before participants made each dilemma judgment, we also assessed how torn they felt between the available response options: *I feel torn between the two answer choices* on a scale from 1 (*completely disagree*) to 7 (*completely agree*; adapted from Jamieson, 1993). We separately computed feeling torn scores regarding incongruent (α = .63) and congruent dilemmas (α = .64).[7]

Results

Conventional dilemma analysis. As before, there was no significant difference in accepting outcome-maximizing harm in classic dilemmas across the distrust (*M* = .56, *SD* = .18) and trust (*M* = .56, *SD* = .19) mindset conditions, *t*(122) = -0.14, *p* = .889, 95% CI_{diff} [-.07, .06]. The Scaled-Information Bayes Factor in favor of the null when scale *r* =.707 was 4.01, suggesting that the data were about 4 times more likely to support the null than alternative hypothesis.

Process dissociation analysis. Once again, we conducted a 2 (mindset: distrust versus trust) × 2 (parameter: utilitarian versus deontological) mixed-model ANOVA with the first factor between-subjects and the second factor within-subjects. Again, the parameters evinced a sensible pattern of correlations with conventional judgments and one another (see Table 1). Consistent with Studies 1 and 2, this analysis revealed no main effect of parameter, $F(1, 122) = .01$, $p = .941$, $\eta_p^2 < .001$, and no interaction, $F(1, 122) = .86$, $p = .356$, $\eta_p^2 = .01$ (see Figure 4). However, we again found the predicted main effect of mindset: Participants in a distrust mindset ($M = .27$, $SD = .66$) scored higher on both parameters than those in a trust mindset ($M = -.33$, $SD = .90$), $F(1, 122) = 18.53$, $p < .001$, $\eta_p^2 = .13$. Post-hoc tests confirmed that the deontology parameter was significantly higher in the distrust ($M = .23$, $SD = .99$) than trust condition, ($M = -.27$, $SD = .95$), $F(1, 122) = 8.44$, $p = .004$, $\eta_p^2 = .07$, 95% CI_{diff} [.16, .86]. This time, the utilitarian parameter was also significantly higher in the distrust ($M = .32$, $SD = .75$) than trust condition ($M = -.38$, $SD = 1.12$), $F(1, 122) = 16.80$, $p < .001$, $\eta_p^2 = .12$, 95% CI_{diff} [.36, 1.03].

Perceived ambivalence. We examined whether feeling torn scores varied across the trust/distrust mindset conditions in response to both incongruent and congruent dilemmas by conducting a 2 (mindset: distrust versus trust) × 2 (dilemma: incongruent versus congruent) mixed-model ANOVA with the first factor between-subjects and the second factor within-subjects. This analysis revealed a significant effect of mindset: overall, participants in the distrust mindset reported feeling more torn ($M = 4.08$, $SD = 1.09$) than participants in the trust mindset ($M = 3.59$, $SD = 1.20$), $F(1, 122) = 5.54$, $p = .020$, $\eta_p^2 = .04$. In addition, there was a main effect of dilemma: overall, considering incongruent dilemmas increased feeling torn ($M = 4.36$, $SD = 1.33$) more than considering congruent dilemmas ($M = 3.34$, $SD =$

1.22), $F(1, 122) = 112.63$, $p < .001$, $\eta_p^2 = .48$. Neither of these effects were qualified by a significant interaction, $F(1, 122) = .02$, $p = .897$, $\eta_p^2 < .001$.

Post-hoc contrasts indicated that people felt significantly more torn between response options for incongruent dilemmas in the distrust ($M = 4.59$, $SD = 1.25$) than trust ($M = 4.11$, $SD = 1.39$) condition, $F(1, 122) = 3.94$, $p = .049$, $\eta_p^2 = .03$, 95% CI_{diff} [.001, .94]. Similarly, people felt significantly more torn regarding congruent dilemmas in the distrust ($M = 3.57$, $SD = 1.18$) than trust ($M = 3.08$, $SD = 1.23$) condition, $F(1, 122) = 5.26$, $p = .024$, $\eta_p^2 = .04$, 95% CIs for the difference [.07, .93]. These findings confirm that a) the distrust mindset indeed increased how torn people felt between the available response options for both kinds of dilemmas, and b) overall, people felt more torn between options for incongruent than congruent dilemmas (consistent with past work on difficulty ratings; Conway & Gawronski, 2013). Moreover, these findings held when controlling for gender, which was important as women typically report more difficulty in resolving such moral dilemmas than do men (Friesdorf et al., 2015).[8]

Deontology parameter mediation. Finally, to determine whether feeling torn between response options mediated the impact of our mindset manipulation on each process dissociation parameter, we used the PROCESS macro in SPSS 22 (IBM Corp., 2013) to conduct two 5000-iteration simultaneous mediation bootstrap analyses according to the procedures recommended by Preacher and Hayes (2004). We coded mindset 1 = *distrust*, 0 = *trust*. The first analysis examined whether feeling torn on each dilemma type simultaneously mediated the impact of distrust (versus trust) mindset on the deontology parameter, controlling for the utilitarian parameter (see Figure 5). Again, as expected, a distrust (versus trust) mindset increased feeling torn on both incongruent dilemmas, $B = .51$, $SE = .25$, $p =$

.042, 95% CI [.02, .99], and congruent dilemmas, B = .75, SE = .21, p < .001, 95% CI [.34, 1.16]. More importantly, there was a significant indirect effect of distrust mindset on the deontology parameter through feeling torn on incongruent dilemmas, B = .11, SE = .09, 95% CI [.001, .35], as the confidence intervals excluded zero. Additionally, there was a significant negative indirect effect on the deontology parameter through feeling torn on congruent dilemmas, B = -.35, SE = .13, 95% CI [-.66, -.14], as these confidence intervals also excluded zero. This pattern held when adding gender as a covariate.

Utilitarian parameter mediation. The second analysis examined whether feeling torn on each dilemma type simultaneously mediated the impact of distrust (versus trust) mindset on the utilitarian parameter, controlling for the deontology parameter (see Figure 6). Again, as expected, we found a significant indirect effect of distrust mindset on the utilitarian parameter through feeling torn on incongruent dilemmas, B = .10, SE = .06, 95% CI [.01, .28], as the confidence intervals excluded zero. Additionally, there was again a significant negative indirect effect through feeling torn on congruent dilemmas, B = -.26, SE = .10, 95% CI [-.50, -.10], as these confidence intervals also excluded zero. Again, this pattern held when adding gender as a covariate.

Discussion

Study 3 replicated the pattern found in Studies 1 and 2: participants in a distrust mindset made similar conventional moral judgments as participants in a trust mindset. However, it appears that the distrust mindset manipulation nonetheless impacted moral dilemma processing: participants in a distrust mindset scored higher on both the deontology and utilitarian parameters than participants in a trust mindset. Moreover, participants in a distrust mindset reported feeling more torn between available answer options than

participants in a trust mindset—and this effect mediated the impact of distrust mindset on the parameters. As incongruent dilemmas pit a deontological option against a utilitarian option, this finding suggests that a distrust mindset induced participants to vacillate between choosing both dilemma options, thereby ramping up the inclination to choose each.

Notably, participants in a distrust mindset also felt torn between response options when facing congruent dilemmas. Congruent dilemmas pit one response option—rejecting harm that does not maximize outcomes (consistent with both deontology and utilitarianism)—against another response option—accepting non-outcome-maximizing harm (inconsistent with both deontology and utilitarianism, but consistent with non-moral motives such as self-interest, vengeance, or sadism). In other words, on congruent dilemmas, distrust increased ambivalence between deontological and utilitarian responses, on the one hand, versus amoral or immoral responses on the other. Hence, ambivalence in congruent dilemmas mediated reductions in both PD parameters—even as ambivalence in incongruent dilemmas simultaneously increased both parameters. This finding indicates that a distrust mindset induced participants to vacillate more between whichever two options were available. In the case of incongruent dilemmas, these two responses reflected utilitarian and deontological concerns respectively, ratcheting up both response tendencies. Conversely, in the case of congruent dilemmas, distrust caused participants to vacillate between one answer that satisfies both deontological and utilitarian concerns versus another answer reflecting amoral or immoral concerns, thereby reducing the tendency to endorse deontological and utilitarian concerns relative to other concerns.

General Discussion

Together, the results of three studies demonstrated a consistent pattern: activating a generalized distrust (versus trust) mindset increased participants' inclination to consider both the deontological and utilitarian response options when facing moral dilemmas. These simultaneous increases cancelled one another out in conventional relative dilemma judgments. Distrust increased reports that people felt torn between the available response options, and feeling torn, in turn, mediated the effect of distrust on both response inclinations. Importantly, the effect of feeling torn on response inclinations depended upon which options people considered. When distrust increased feeling torn on incongruent dilemmas where rejecting harm is consistent with deontology and accepting harm is consistent with utilitarianism, people scored higher on both deontological and utilitarian response inclinations. Conversely, when distrust increased feeling torn on congruent dilemmas where rejecting harm is consistent with both deontology and utilitarianism, and accepting harm is consistent with non-moral motives, people scored lower on both deontological and utilitarian response inclinations. However, people felt more torn between response options for incongruent than congruent dilemmas overall, leading to an overall effect of distrust increasing both deontological and utilitarian response inclinations across three studies.

This pattern suggests that upon encountering a moral dilemma where people prefer one or the other response by default, distrust jostled people out of their default response strategy, inducing them to more carefully consider the available alternative response, ratcheting up their desire to select both responses. Conversely, trust allowed people to uncritically select whichever response they prefer by default without carefully considering the alternative, resulting in net lower response inclinations overall. Such an effect is invisible

when examining judgments of only classic dilemmas that treat deontological and utilitarian responses as opposites, because these simultaneous increases in response inclinations cancel one another out.

Note that not all findings reached significance (specifically in Study 2), and we conducted three additional studies that remained in the file drawer. Therefore, we conducted both fixed and random effect meta-analyses across all samples, correcting for differences in sample sizes (Hunter & Schmit, 1990). Fixed effect analysis revealed significant effects for both the deontology, ES = -.20, *SE* = .03, *Z* = -5.93, $p < .001$, 95% CI [-.27, -.14], and utilitarian parameters, ES = -.14, *SE* = .03, *Z* = -4.08, $p < .001$, 95% CI [-.21, -.07]. However, the random effects model revealed a significant effect only for the deontology parameter, ES = -.24, *SE* = .11, *Z* = -2.23, $p = .026$, 95% CI [-.45, -.03], as the effect for the utilitarian parameter fell below threshold, ES = -.20, *SE* = .14, *Z* = -1.36, $p = .173$, 95% CI [-.48, .08] (see Supplement for details).

Implications for Trust/Distrust and Moral Judgment Research

These findings mesh with the broader literature on distrust mindsets, which suggest that distrust increases the activation, generation, and selection of creative alternatives and alternatives to default positions (e.g., Mayer & Mussweiler, 2012; Posten & Mussweiler, 2013; Schul, et al., 2004). In the current work, generalized distrust increased reports of feeling torn between whichever choice options were presented, and participants in a distrust mindset evinced stronger inclinations to select both options. This pattern corroborates past work on generalized trust and distrust mindsets indicating that distrust leads people to engage in non-routine processing where they carefully consider alternative options rather than uncritically accepting their initial perspective (e.g., Posten & Mussweiler, 2013). Hence,

these findings provide further support to cognitive models of generalized trust and distrust while extending investigation of this phenomenon from decision-making (Schul & Peri, 2015) to moral decision-making in particular.

Moreover, these findings are inconsistent with other plausible predictions. If generalized undermines confidence in evidence, similar to distrust of specific targets (e.g., Darke & Ritchie, 2007; Schul & Peri, 2015), then distrust may have reduced one or both parameters, by reducing confidence that actions will cause harm or lead to beneficial outcomes. If generalized distrust operates by increasing negative affect, it should have selectively increased the deontological parameter (Conway & Gawronski, 2013), and if it operated by inducing a general negativity bias then it should have increased the acceptability of harm, hence reducing both parameters. Yet, instead distrust increased both the deontology and utilitarian parameters, supporting our argument that distrust mindsets ratchet people up between response alternatives. In addition, our findings have implications for moral dilemma research. Specifically, these findings contribute to a growing body of work that illustrates how conventional data analytic techniques fail to fully describe the psychological processes involved in moral decision-making. Conventional methods assess only the relative contributions of processes underlying dilemma judgments. Had we conducted these studies using conventional relative judgments, we would have concluded that inducing a distrust mindset has no impact on the processes underlying dilemma judgments because these relative judgments themselves did not change. In contrast, employing process dissociation to independently assess the absolute strength of each underlying process enabled us to detect how distrust simultaneously increased both deontological and utilitarian inclinations relative to trust. This pattern resembles the relation between moral identity and moral judgments:

moral identity positively correlated with both parameters, which canceled out in conventional judgments (Conway & Gawronski, 2013). Such results are of paramount importance in light of recent calls to abandon moral dilemma research due to confusing relations between conventional judgments and other measures (Kahane, 2015). The current findings suggest that such calls are premature, as confused relations may be the product of conventional data analytic techniques rather than conceptual errors. Employing process dissociation may clarify these otherwise confusing findings.

Quality of Moral Decision-Making

Our findings suggest that when facing difficult ethical decisions, generalized distrust mindsets may lead people to more carefully consider the available options before ultimately arriving at a given judgment. Therefore, distrust mindsets have the potential to improve the quality of moral decision-making—depending on which options people consider. If people consider a choice between two morally-defensible alternatives, as in classic dilemmas where causing harm maximizes outcomes, then distrust may induce them to engage in a more balanced consideration of both alternatives rather than rushing to uncritically select their default preference. Granted, it is possible for people to arrive at the same dilemma decision using different strategies, but it seems reasonable that decisions based on considering both sides of complex moral issues, rather than zealous focus on a single aspect, are usually preferable, despite the subjectively negative experience of decision conflict. Consistent with this perspective, considering multiple aspects of controversial issues can be socially adaptive (Pillaud, Cavazza, & Butera, 2013); the way people arrive at moral judgments can have important social consequences (e.g., Critcher, Inbar, & Pizarro, 2013). Hence, distrust may be effective at improving decisions where there are moral merits to both sides, such as

deciding which friend to support in a disagreement where both are partially at fault, or settling land disputes between social groups who both have legitimate claims. Considering both options may allow individuals to take the perspective of others, which can be important in conflicts about moral issues.

However, distrust might be problematic when people face simpler decisions where one option is unethical but tempting and the other is ethical. Under distrust, considering the ethical option should also automatically activate the unethical alternative, increasing temptation to select it, and increased decisional conflict. Such cases may include congruent moral dilemmas, but also decisions such as whether to accept a bribe or whether or not to cooperate in social dilemmas. In such cases, trust mindsets may be more effective at promoting ethical decision-making, especially considering that many people's initial inclinations tend towards the prosocial (Rand, Greene, & Nowak, 2012). Future work should consider clarifying the pragmatic potential of applying distrust and trust mindsets to improve decision-making in different contexts.

Specific versus Generalized Distrust

In the current work, we examined the impact of generalized distrust and trust mindsets, where the effects of trust/distrust cues extent to information and targets beyond the distrust-eliciting stimulus (consistent with a large body of work, e.g., Schul, et al., 2004). Note that the potentially positive impact of generalized distrust on moral decision-making stands in opposition to the impact of interpersonal distrust towards a specific target, which can reduce the ethicality of moral decisions as people act selfishly due to fear of being exploited (Gollwitzer, Rothmund, & Süßenbach, 2013). Hence, generalized distrust and particularized distrust of specific individuals may have very different impacts on moral

judgment. Whereas target-specific distrust may evoke the same general cognitive processes as generalized distrust (i.e., non-routine processing), these effects may be superimposed by the more specific—and motivational—impacts of specific interpersonal distrust (e.g., fear of exploitation, suspicion, negative evaluations, concern about malevolent intentions), which may result in reduced ethical behavior. This idea is consistent with previous work on the differential effects of "diffuse" and "focused" trust/distrust in the domain of decision making (Schul & Peri, 2015).

Limitations

Although the current findings provide new insight into the relation between distrust and moral judgment, like all studies they suffer from some limitations. First, it remains unclear from the current work whether the impact of distrust on moral judgment is due to non-social aspects of distrust (e.g., distrusting one's initial inclination, as advised by Charles de Talleyrand) or social aspects of distrust (e.g., distrusting particular interaction partners, as advised by Frank Herbert). We have framed our argument in terms of the cognitive elements of distrust, due to the unfocused and target-independent way we activated trust and distrust mindsets in our studies. Hence, the cognitive aspects of distrust seem to be the most plausible explanation for the effects, in line with the existing literature on trust and distrust mindsets (e.g., Schul et al., 2004, 2008), and the dual-process model of moral judgment (Greene et al., 2001). However, in addition, recent work has begun to explore the social implications of moral judgments, and suggests that people link dilemma judgments to the moral self (Conway & Gawronski, 2013), and modify their judgments depending on social circumstances (Lucas & Livingston, 2014). Therefore, social aspects of distrust may also

play a role in moral judgments. Future work might profitably examine this by separately inducing social and non-social distrust mindsets.

Another limitation of the present work is that mood effects cannot be ruled out completely, as we did not assess this construct. However, it seems implausible that mood could explain the results. Neither Mayer and Mussweiler (2011), nor Posten and Mussweiler (2013), nor Mayo and colleagues (2014) found any effects on mood using similar paradigms. Moreover, dilemma research suggests that negative emotions push dilemma judgments in specific directions (Strohminger, Lewis, & Meyer, 2011; Valdesolo & DeSteno, 2006)—that is, selectively increasing deontological responses (Conway & Gawronski, 2013)—rather than simultaneously increasing both response tendencies as seen in the current data. Thus, although we cannot completely rule out the possibility that mood influenced effects in the current work, such an explanation would be inconsistent with past findings.

These findings also leave an important theoretical question regarding the mechanism involved. On the one hand, distrust may increase the *salience* of alternative responses—such that responses which would have been discarded previously now enters consideration. In other words, distrust reduces the threshold needed for consideration of alternatives. On the other hand, distrust may increase the *weight* assigned to the alternative response. In this view, people typically experience a modest degree of vacillation between responses, but distrust increases the weight of each consideration, thereby increasing the degree of vacillation. Current results are potentially compatible with either mechanism. Future work should seek to distinguish between them. The analyses of classic dilemma responses and process dissociation parameters bear another caveat, relying on aggregate data and a between-subjects design. Specifically, even though trust versus distrust mindsets did not evince an

effect on overt dilemma judgments, it may potentially be the case that distrust affected individual participants in different directions, cancelling out in the overall sample. Such a possibility would be in line with some participants, potentially those with a predominantly utilitarian default response, showing a greater increase in the deontology parameter and others showing a relatively stronger increase in the utilitarian parameter. While we assume that the cognitive processes involved in distrust should lead to an increase in both judgment inclinations within participants, such a pattern may also possibly produce similar results. Therefore, while present studies robustly show that distrust versus trust does affect moral processing via increased ambivalence, the possibility remains that distrust has other effects. For example, recent work suggests that dilemma decisions may involve an element of self-presentation (Rom, Weiss, & Conway, 2017). It remains possible that distrust makes people more acutely aware of self-presentational concerns, thereby altering judgments. Future work should examine this possibility.

Finally, the current work suffers from a limitation common to all moral dilemma research: moral dilemmas are (necessarily) hypothetical scenarios (due to the extreme harm involved) and therefore may not always reflect people's actual decisions in real life (Francis et al., 2016; Patil, Cogoni, Zangrando, Chittaro, & Silani, 2014). Although it is certainly desirable to employ ecologically valid stimuli, nonetheless hypothetical stimuli remain valuable if they allow for drawing distinctions between different psychological processes and demonstrate how each process contributes to judgment, as moral dilemmas do (Cushman & Greene, 2012). In real life, perhaps the contributions of these processes may differ, but that does not alter the utility of understanding the operation of each contributing process, as the current paradigm allows. Moreover, there are numerous real-life examples of moral

dilemmas. For example, the German Parliament and courts recently debated whether it is acceptable for the military to destroy a hijacked civilian airliner to prevent it from causing widespread destruction (Whitlock, 2006), and the acceptability of harm to innocent civilians ('collateral damage') is a common consideration in military interventions aimed at stabilizing collapsing regimes. Therefore, moral dilemmas are less hypothetical than many people may assume.

Conclusion

Returning to the animal research dilemma described at the beginning—should you torture animals to help cure AIDS? Our findings suggest that if you experience a distrustful state of mind, you are more likely to question your default inclination and consider the alternative response. On the other hand, you may uncritically select your default dilemma response if you are in a trusting state of mind. But beware! Frank Herbert and Maurice de Talleyrand were skeptical of those who fail to question their initial position, even (or especially) when trying to improve the world. Many ethical problems involve complex trade-offs between different parties. The present research suggests that generalized distrust may result in more balanced consideration of these different interests in complex moral decision-making. Sometimes it may be wise to distrust one's moral compass.

Acknowledgement

This research was funded by the German Research Foundation (DFG) through a Gottfried Wilhelm Leibniz Award awarded to Thomas Mussweiler (MU 1500/5–1) as well as a research grant within the research unit Design and Behavior awarded to Thomas Mussweiler und Axel Ockenfels (MU 1500/7-2).

Notes

[1] Note that the influence of distrust is qualitatively different than merely inducing greater cognitive effort. Distrust should jostle people out of their default response strategy regardless of which response they preferred by default. Conversely, inducing cognitive effort should selectively increase the utilitarian parameter, rather than both parameters (Conway & Gawronski, 2013).

[2] Note that this application of PD does not allow for the calculation of a 'guessing parameter' to disentangle construct-irrelevant responding (e.g., inattentive responding) from construct-relevant responding. Unlike other applications of PD (Payne & Bishara, 2009), there is only one congruent and incongruent block, resulting in insufficient degrees of freedom for a guessing parameter to vary independently. Hence, this PD model requires making one of two assumptions in the case where neither utilitarian nor deontological concerns drive dilemma judgments, depending on which calculation model one employs. Here we employ the 'U-dominant' model, which begins with deriving utilitarian inclinations, and assumes that only when such inclinations do not drive responses, then deontological inclinations may drive responses. In PD, the third case (when neither utilitarian nor deontological concerns drives responses) is mathematically constrained to be the opposite of the second case, so the U-dominant model makes the assumption that in the absence of U or D, participants will systematically accept harm (see Figure 1). We suggest this assumption is not unreasonable theoretically; it suggests that in the absence of deontological or utilitarian concerns, people feel free to accept harm. Yet, this assumption will be inaccurate to the degree that construct-irrelevant responding leads participants to reject harm (e.g., inattentive responding may result in a roughly equal mixture of harm acceptance and rejection). To the degree that this

assumption is violated, the parameters may be inflated because construct-irrelevant responding that leads people to reject harm will load on the D parameter, and will also load on the U parameter when that harm-rejection occurs in congruent dilemmas. Clearly, this assumption is suboptimal; yet, we argue it remains preferable to the assumptions made by the alternative 'D-dominant' model. The D-dominant model starts by deriving deontological inclinations (always reject harm), and assumes that only when such inclinations do not drive responses then utilitarian inclinations may do so. Again, the third case is constrained to be the opposite of the second case, so this model makes the less-plausible theoretical assumption that in the absence of deontological or utilitarian concerns, participants will demonstrate an 'anti-utilitarian' pattern of 'maximum carnage:' only cause harm that makes the world worse, and never cause harm that makes the world better. This assumption has the advantage of including both harm acceptance (for congruent dilemmas) and harm rejection (for incongruent dilemmas) in the third case, which suggests that non-construct variance will somewhat cancel out to the degree that it entails randomly rejecting vs. accepting harm. However, to the degree that construct-irrelevant responding leads participants to accept harm for incongruent dilemmas and reject harm for congruent dilemmas, this variance will inflate the U parameter.

There are both theoretical and empirical reasons to prefer the U-dominant model (see Conway and Gawronski, 2013, Appendix B). Most important, it seems more theoretically plausible that in the absence of either deontological or utilitarian concerns (neither concern for causing harm nor maximizing outcomes), people evince a general, non-specific disregard for others' well-being, which also enables them to pursue other goals, such as self-interest (for example, in the congruent crying baby dilemma, causing harm allows one to escape

working in a mine). We find this assumption more plausible than the D-dominant assumption that in the absence of either deontological or utilitarian concerns, people evince a response strategy that systematically selects the worst possible outcome (for example, this would involve killing the baby to avoid working in a mine, but refusing to kill the baby to save oneself from death). Moreover, by beginning with deriving the process theorized to involve more cognitive deliberation, the U-dominant model is analogous to all other PD research, which also begin by deriving parameters tapping controlled processing (Payne & Bishara, 2009), except one paper on Stroop performance (Lindsay & Jacoby, 1994). Therefore, we suggest the U-dominant model, despite making suboptimal assumptions, nonetheless remains preferable to the D-dominant model with its even more suboptimal assumptions.

One way to partially address concerns over these problematic assumptions in the absence of a guessing parameter is to compute both the U-dominant and D-dominant parameters and see whether effects hold across both calculations. Indeed, the main effect of the trust mindset manipulation on the D-dominant parameters demonstrated a very similar pattern as with U-dominant parameters: Study 1, $F(1, 86) = 6.12$, $p = .015$, $\eta_p^2 = .07$, Study 2, $F(2, 158) = 2.58$, $p = .079$, $\eta_p^2 = .032$, Study 3, $F(1, 131) = 24.25$, $p < .001$, $\eta_p^2 = .16$ (note the D-dominant method requires excluding a few participants due to division-by zero errors). Moreover, the D-dominant parameters correlated highly with their U-dominant counterparts. As these methods employ somewhat different calculations and make different assumptions about the pattern of responding in the third case, the fact that patterns hold across both models suggests that concerns about violated assumptions may not be so great as to threaten construct validity.

Note that some theorists have aimed to tackle this issue by including additional dilemmas where action is coded as leading to a deontological decision and inaction is coded as leading to a utilitarian decision: the CNI Model (Gawronski et al., 2015; 2017; 2018). This model allows for estimation of three parameters: a *C parameter* tracking adherence to consequences, an *N parameter* is designed to track consistency with deontological norms, and an *I parameter* tracking systematic inaction tendencies to either act or refrain from action independent of consequences or deontological norms. Although this model suggests a promising avenue for disentangling systematic but construct-irrelevant responding from theoretically relevant responding, in doing so it fundamentally changes the nature of the constructs under examination as represented by the parameters of interest. Specifically, in classic dilemmas the deontological option always requires refraining from causing harm; in contrast, in reversed CNI dilemmas the deontological option requires acting to save one person at the expense of inflicting suffering on others. Although this action is coded as 'consistent with deontological norms,' it remains unclear whether this coding best describes this action. Both deontological theorists (e.g., Kant, 1785/1959; Baron, 1994) and lay people (e.g., Cushman, Young, & Hauser, 2006; Janoff-Bulman, Sheikh, & Hepp, 2009) endorse the distinction between causing harm via action versus omission as an important deontological distinction, yet the CNI model equates these two forms of action in their capacity to represent the adherence to deontological norms. Moreover, this action causes harm, suggesting it should also be coded as inconsistent with deontological norms, as any action that causes harm violates deontological norms against causing harm. Indeed, it may be impossible to present scenarios where action leads to a decision truly consistent with deontological

norms—deontological norms may always require refraining from action. Hence, we recommend caution when comparing the results of PD studies with the CNI model.

[3] Results of Study 1 when excluding three people were nearly identical to the full sample. Conventional dilemmas were not significantly impacted by the manipulation, $t(83) = -0.01$, $p = .996$, 95% CI_{diff} [0-.09 .09], but the main effect of condition on the parameters remained significant, $F(1, 83) = 6.47$, $p = .013$, $\eta_p^{2 =} .072$. Post-hoc tests confirmed that the deontology parameter remained significant, $F(1, 83) = 4.35$, $p = .040$, $\eta_p^2 = .050$, 95% CI_{diff} [.02, .89], and the utilitarian parameter slightly improved but remained marginal, $F(1, 83) = 3.39$, $p = .069$, $\eta_p^2 = .039$, 95% CI_{diff} [-.03, .84].

[4] A recent meta-analysis indicated that gender is a strong predictor of moral dilemma judgments, with women scoring much higher than men on the deontology parameter, and men scoring slightly higher than women on the utilitarian parameter (Friesdorf et al., 2015). The pattern of gender effects in all of the current studies matched this meta-analytic pattern. Therefore, in all studies we re-ran the main analysis treating gender as a covariate. In Study 2, the omnibus effect of condition weakened when gender was included as a covariate to $F(2, 160) = 2.28$, $p = .107$. In all other cases, the effects of trust/distrust mindsets remained significant when controlling for gender. Hence, the current findings cannot be attributed to gender. [5] At the end of the study, participants also reported perceived task difficulty for the scrambled sentences (*How difficult did you find the task where you unscrambled the sentences?* 1 = *very easy*, 7 = *very difficult*), and dilemmas (*How difficult did you find it to make a decision for the stories overall?* 1 = *very easy*, 7 = *very difficult*), as well as dilemma decision importance (*How important was it to you to make the right decision for the stories overall?* 1 = *not at all important*, 7 = *very important*). We assessed these questions to

determine whether overall perceptions of the tasks varied across condition. Task difficulty for the scrambled sentences did not significantly vary across the distrust (M = 2.76, SD = 1.58), neutral (M = 2.33, SD = 1.35), and trust (M = 2.22, SD = 1.32) mindset conditions, $F(2, 161) = 2.12$, $p = .123$, $\eta_p^2 = .03$. Nor did dilemma difficulty perceptions significantly vary across the distrust (M = 3.53, SD = 1.70), neutral (M = 3.38, SD = 1.78), and trust (M = 3.83, SD = 1.74) mindset conditions, $F(2, 161) = .97$, $p = .383$, $\eta_p^2 = .01$. Unexpectedly, perceived decision importance did vary significantly across the distrust (M = 6.12, SD = .88), neutral (M = 6.52, SD = .70), and trust (M = 6.16, SD = .99) mindset conditions, $F(2, 161) = 3.33$, $p = .038$, $\eta_p^2 = .04$. Post-hoc comparisons indicated that participants in the neutral condition rated their dilemma decisions as more important than participants in either the distrust, $t = 2.35$, $p = .024$, $d = 0.50$, 95% CI_{diff} [-.74, -.05], or trust conditions, $t = 2.25$, $p = .029$, $d = 0.42$, 95% CI_{diff} [-.68, -.04], which did not differ from one another, $t = -0.25$, $p = .828$, $d = -0.04$, 95% CI_{diff} [-.36, .29]. However, adding this variable as a covariate to the main analysis does not appreciably change the parameter pattern, so this finding cannot explain the effects of distrust mindsets on the parameters.

[6] Note that in Study 2, perceived dilemma difficulty did not vary across condition, but we measured it only once at the end of the study. Thus, it reflects a recollection of perceived difficulty across the entire battery of moral dilemmas after all decisions are complete. Conversely, in Study 3 we measured online perceptions of feeling torn between available response alternatives for each dilemma during the moment of deliberation before participants reported their dilemma judgment. Hence, although these measures are conceptually similar, we expected the richer online measure in employed in Study 3 to demonstrate sensitivity to the mindset manipulation even though dilemma difficulty in Study 2 did not.

[7] Although reliabilities for the measures of feeling torn were somewhat low, this is to be expected for dilemmas representing a wide range of variation in content. Note that lower than ideal reliabilities make the reported analyses more conservative by increasing the difficulty of finding significant effects.

[8] In this study we again assessed perceptions of the task difficulty for the scrambled-sentences task. Difficulty perceptions were similar across the distrust (M = 2.37, SD = 1.52) and trust (M = 2.46, SD = 2.08) mindset conditions, $t(132) = -0.30$, $p = .764$, $d = -0.05$, 95% CI_{diff} [-.54, .68].

References

Amazon (2015). https://requester.mturk.com/ Retrieved on April 1, 2015.

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, 23, 861-868. doi:10.1177/0956797611434965

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

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*, 381–417. doi:10.1016/j.cognition.2008.03.001

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences*, *17*, 1-10.

Baron, J., & Gürçay, B. (2016). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition, 4*, 556-575.

Burnes, B. (2003). *Harry S. Truman: His life and times*. Kansas City, MO: Kansas City Star Books.

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*, 3-5. doi:10.1177/1745691610393980

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*, 216-235. doi:10.1037/a0031021

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*, 308-315.

Cushman, F. A., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience,* 7, 269-279. doi:10.1080/17470919.2011.614000

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological science*, *17*, 1082-1089.

Darke, P. R., & Ritchie, R. J. (2007). The defensive consumer: Advertising deception, defensive processing, and distrust. *Journal of Marketing Research, 44*, 114-127.

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.

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*, 1164-1184.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5*, 5-15.

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual Morality: Transitioning from Moral Judgment to Moral Action? *PLoS ONE, 11*, e0164374.

Friesen, J., & Sinclair, L. (2011). Distrust and simultaneous activation of multiple categories. *Social Psychological and Personality Science*, *2*, 112-118.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin, 42*, 696-713. doi:10.1177/0146167215575731

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*, 415-426.

Greene, J. D. (2015). The rise of moral cognition. *Cognition, 135*, 39-42.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108.

Helzer, E. G., Fleeson, W., Furr, R. M., Mendl, P., & Barranti, M. (2016). Once a utilitarian, consistently a utilitarian? Examining principledness in moral judgment via the robustness of individual differences. *Journal of Personality*. doi:10.1111/jopy.12256

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Jamieson, D. W. (1993, August). *The attitude ambivalence construct: Validity, utility, and measurement.* Paper presented at the annual meeting of the American Psychological Association, Toronto.

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: two faces of moral regulation. *Journal of personality and social psychology*, *96*, 521.Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell

us little (or nothing) about utilitarian judgment. *Social Neuroscience*. doi:10.1080/17470919.2015.1023400

Kant, I. (1785/1959). *Foundation of the Metaphysics of Morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.

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.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature 446*, 908-911. doi:10.1038/nature05631

Lee, J. J., & Gino, F. (2015). Poker-faced morality: Concealing emotions leads to utilitarian decision making. *Organizational Behavior and Human Decision Processes, 126*, 49-64.

Lucas, B. L., & Livingstone, R. W. (2014). Feeling socially connected increases utilitarian choices in moral dilemmas. *Journal of Experimental Social Psychology, 53*, 1–4. doi:10.1016/j.jesp.2014.01.011

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

Mayo, R., Alfasi, D., & Schwarz, N. (2014). Distrust and the positive test heuristic: Dispositional and situated social distrust improves performance on the Wason Rule Discovery Task. *Journal of Experimental Psychology: General, 143*, 985.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *TRENDS in Cognitive Sciences, 11*, 143-152. doi:10.1016/j.tics.2006.12.007

Mill, J. S. (1861/1998). *Utilitarianism*. In R. Crisp (Ed.), New York: Oxford University Press.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience, 9*, 94-107.

Park, G., Kappes, A., Rho, Y., & Van Bavel, J. (2016). At the heart of morality lies neuro-visceral integration: lower cardiac vagal tone predicts utilitarian moral judgment. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nsw077

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81,* 181–192.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology, 20*, 272-314. doi:10.1080/10463280903162177

Pillaud, V., Cavazza, N., & Butera, F. (2013). The social value of being ambivalent self-presentational concerns in the expression of attitudinal ambivalence. *Personality and Social Psychology Bulletin*, *39(9)*, 1139-1151

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. doi:10.1037/a0033170

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717-731.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*, 427-430.

Rom, S., 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.

Rom, S., & Conway, P. (2018). The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Psychology*.

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, 32*, 228–253. doi: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.

Schul, Y., Mayo, R., & Burnstein, E. (2008). The value of distrust. *Journal of Experimental Social Psychology*, *44*, 1293-1302.

Schul, Y., & Peri, N. (2015). Influences of distrust (and trust) on decision making. *Social Cognition*, *33*, 414–435. http://doi.org/http://dx.doi.org/101521soco2015335414

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Strohminger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition, 119*, 295–300. doi:10.1016/j.cognition.2010.12.012

van Harreveld, F., Rutjens, B. T., Rotteveel, M., Nordgren, L. F., & Van Der Pligt, J. (2009). Ambivalence and decisional conflict as a cause of psychological discomfort: Feeling tense before jumping off the fence. *Journal of Experimental Social Psychology, 45*, 167-173. doi:10.1016/j.jesp.2008.08.015

Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*, 476–477. doi:10.1111/j.1467-9280.2006.01731.x

Whitlock, C. (2006, February 16). German court overturns law allowing hijacked airliners to be shot down. *Washington Post Foreign Service*. Retrieved from http://www.washingtonpost.com

Appendix A– Scrambled-Sentences Distrust/Trust Mindset Induction

Dilemma and Scrambled Sentence Task Instructions:

On the following screens you will see a series of **20** short stories. Please read them carefully, then make a judgment about the stories. There is no right or wrong answer; we are simply interested in your personal opinion. Even though some stories may seem similar, each story is different in important ways. Please note that some stories refer to things that may seem unpleasant to think about. This is because we are interested in people's thoughts about difficult, real-life issues.

In between the stories, we will also ask you to complete a language skill task. More precisely, after each of the stories you will be asked to create a sentence from a set of words, twice. Each set consists of five words, but you will be asked to use only four of these words to form a *grammatically correct* sentence. Please use only the words that are provided without changing any of the words by adding or subtracting letters. Uppercase and lowercase are not important.

Sometimes, there may be several solutions to this task. Any grammatically correct sentence is fine. Please write down the sentence you created in the space provided, and work as fast and accurately as possible.

Example: GROW TREES BAG ON LEMONS Solution: Lemons grow on trees.

Before the first story, we will give you ten practice scrambled sentences to get used to the task. Please click on "Continue" to proceed to the first practice sentence.

Distrust Condition

Interspersed with neutral sentences.

Scrambled Sentence	Solution
UNRELIABLE SOME ARE BAG FRIENDS	Some friends are unreliable
TWO-FACED A BIRD PERSON HE'S	He's a two-faced person
FELT DISBELIEF SOMETIMES PEN SHE	Sometimes she felt disbelief
APPEARS PATH THE RAKES UNCERTAIN	The path appears uncertain
PLANS UNDERMINED THE CEREAL THEY	They undermined the plans
THE IS DISTRUSTED PEACE HE	He distrusted the peace
FAKED COUCH TEST WAS THE	The test was faked
TRUST THEY THEM POOL CAN'T	They can't trust them
QUESTIONES DRINK ASKED HE MISLEADING	He asked misleading questions
SUSPICION THERE MAN GROWING WAS	There was growing suspicion
ARTICLE WAS THE FABRICATED DOOR	The article was fabricated
A ANSWER SLIPPERY GOING THAT'S	That's a slippery answer
LEAF IS UNSOUND ARGUMENT THE	The argument is unsound
ULTERIOR HAS RUG HE MOTIVES	He has ulterior motives
DOUBLE A AGENT HE'S UMBRELLA	He's a double agent
BETRAYAL REEKS OF STAPLER THAT	That reeks of betrayal
HE'S MILK SHADY A CHARACTER	He's a shady character
THOUGHT PLACING THAT'S TREACHEROUS A	That's a treacherous thought
LYING THEY'RE WINDOW HER TO	They're lying to her
A BUG DECEPTIVE IT'S PLAN	It's a deceptive plan
DISTRUSTFUL A CAR CHARACTER SHE'S	She's a distrustful character
MESSAGE DOUBTS THE COFFEE HE	He doubts the message
DISGUISE THEY BOTTLE IN APPEAR	They appear in disguise
VERY SMELLED COOKING IT FISHY	It smelled very fishy

Trust Condition

Interspersed with neutral sentences.

Scrambled Sentence	Solution
RELIABLY CAR STARTS THE BALCONY	The car starts reliably
ALWAYS CHERRY HONESTY BEST IS	Honesty is always best
DOCUMENT THE AUTHENTIC IS SING	The document is authentic
INTEGRITY WITH BEHAVED MARMELADE THEY	They behaved with integrity
REWARDED TRUST BOOTH WAS HER	Her trust was rewarded
UPRIGHT AN HE'S GLASS CITIZEN	He's an upright citizen
UNANIMOUS WAS DEER VOTE THE	The vote was unanimous
HER HE SING ON RELIES	He relies on her
SPOTLESS CHARACTER RUNNING IS HIS	His character is spotless
STORY HER BELIEVABLE APPLE WAS	Her story was believable
RIVER INNOCENT THEIR WAS BEHAVIOR	Their behavior was innocent

IN CUP HE HER CONFIDES	He confides in her
PROMISE THE LION KEPT WAS	The promise was kept
UNDENIABLE AN GUITAR IT'S FACT	It's an undeniable fact
STATEMENT ACCURATE WAS CLOUD HIS	His statement was accurate
IS BIKE COMPLETELY SHE TRUSTWORTHY	She is completely trustworthy
APPEARS PATH THE RAINCOAT CERTAIN	The path appears certain
THEM BELIEVES MORNING SHE IN	She believes in them
ARGUMENT BULLETPROOF A IT'S LAKE	It's a bulletproof argument
DEPENDABLE IS SHE LAMP VERY	She is very dependable
FRIENDS THEY CUSHION LOYAL ARE	They are loyal friends
FAITHFUL PANTS HE'S HUSBAND A	He's a faithful husband
TOLD SHE THE BUG TRUTH	She told the truth
A HE FEATHER TEAMPLAYER IS	He is a teamplayer

Control Sentences (All Studies)
Interspersed with trust and distrust sentences.

Scrambled Sentence	Solution
WATERS BEEN PLANTS SHE THE	She waters the plants
DRINK COLOURS ARE THREE THERE	There are three colours
WAS YESTERDAY DOOR RAINING IT	It was raining yesterday
THEY HANDLE COFFEE THEIR DRINK	They drink their coffee
CAR RAVEN DROVE HE HIS	He drove his car
OFF TURN DISTRIBUTE OFF THE	Turn off the light
PASS CLOUDS BACKWARDS BY THE	The clouds pass by
COOKIES HAT SOME THEY BAKE	They bake some cookies
IT THE SUNNY TODAY IS	It is sunny today
HAVE IS THE ROUND TABLE	The table is round
FRESH THERE QUICKLY BEANS ARE	There are fresh beans
NOW IS WINTER CLICK IS	Now it is winter
PRODUCTS BRUSHES HAIR SHE HER	She brushes her hair
DOG WALKS CORN HE HIS	He walks his dog
THERE WANT MORE WERE FIVE	There were five more
THE BLOWING POPCORN IS WIND	The wind is blowing
THEY WALKING HER ORANGE SAW	They saw her walking
ARE HONEY LEAVES THE FALLING	The leaves are falling
SINGING HE ALONG IS SAUSAGE	He is singing along
CHOCOLATE LIKE THEY COOK EATING	They like eating chocolate
IS GRASS THE BUTTER GROWING	The grass is growing
IS BURNING CANDLE THE STONE	The candle is burning
FIDDLE SEAT PLAYED SHE THE	She played the fiddle
WATER BOOK THEY THE READ	They read the book

Additional Control Sentences (Study 2 neutral condition)
Replaced trust and distrust sentences in Study 2 control condition.

Scrambled Sentence	Solution
COZY BLANKET IS TABLE THE	The blanket is cozy
THOUGHT AN MONDAY THAT'S INTERESTING	That's an interesting thought
WOOD IS CARVES THE SHE	She carves the wood
THE LOOKS ALWAYS BIG TOWER	The tower looks big
WHEN USEFUL VERY IS ELECTRICITY	Electricity is very useful
EARLY MEETING SUITCASE STARTED THE	The meeting started early
DISHWASHER THE RUNNING MOUNTAIN WAS	The dishwasher was running
BOUGHT SHE A MILK CD	She bought a CD
HOMEWORK LIGHTBULB HE HIS BEGAN	He began his homework
DOES HERE HE LAUNDY THE	He does the laundry
LIPSTICK ANOTHER SHE APPLIES HER	She applies her lipstick
IS THE DRYING PAINTING PRODUCTS	The painting is drying
HAY PHONE HORSE THE ATE	The horse ate hay
PENCIL A VISITS HE MUSEUM	He visits a museum
SOCCER NOW PLAY GREY THEY	They play soccer now
APPLE CHAIR ROCKED HIS HE	He rocked his chair
TULIPS BOX THE YELLOW ARE	The tulips are yellow
WEARS SHIRT SCREEN A HE	He wears a shirt
HANDLE PLUS TURNED THEY THE	They turned the handle
TOMATOES FURRY SALE ON ARE	Tomatoes are on sale
WRITE THEIR PAPERS THEY PURSE	They write their papers
BIKE HE THE SPOKE RIDES	He rides the bike
BOTTLED OLD SEEMS RUG THE	The rug seems old
HE BOX SENDING THE EMPTIES	He empties the box

Figure Captions

Figure 1.

Processing tree illustrating the underlying components leading to judgments that harmful action is either acceptable or unacceptable in congruent and incongruent moral dilemmas.

Figure 2.

Mean standardized PD utilitarianism and deontology scores in the trust and distrust mindset conditions, Study 1. Error bars depict standard errors.

Figure 3.

Mean standardized PD utilitarianism and deontology scores in the trust, neutral, and distrust mindset conditions, Study 2. Error bars depict standard errors.

Figure 4.

Mean standardized PD utilitarianism and deontology scores in the trust and distrust mindset conditions, Study 3. Error bars depict standard errors.

Figure 5.

Feeling torn between the response options for incongruent dilemmas mediated the effect of distrust mindset on the deontological parameter, controlling for the utilitarian parameter, Study 3. Feeling torn between the options for congruent dilemmas also simultaneously mediated this effect, but in the opposite direction. Unbracketed values reflect unstandardized

coefficients; bracketed values reflect standard errors. † p = .062, * = p < .05, ** = p < .01, *** = p < .001.

Figure 6.

Feeling torn between the response options for incongruent dilemmas mediated the effect of distrust mindset on the utilitarian parameter, controlling for the deontology parameter, Study 3. Feeling torn between the options for congruent dilemmas also simultaneously mediated this effect, but in the opposite direction. Unbracketed values reflect unstandardized coefficients; bracketed values reflect standard errors. * = p < .05, ** = p < .01, *** = p < .001.

Table 1.

Correlations between Conventional Relative Utilitarian versus Deontological Judgments, the Deontology and Utilitarianism Process Dissociation Parameters, and Gender, All Studies.

	Relative Utilitarian versus Deontological Judgments	**Utilitarian PD Parameter**	**Deontology PD Parameter**
Study 1			
Utilitarian PD Parameter	.49***		
Deontology PD Parameter	-.69***	.22*	
Gender (m = 1, f = 2)	-.21†	-.04	.18
Study 2			
Utilitarian PD Parameter	.48***		
Deontology PD Parameter	-.74***	.18*	
Gender (m = 1, f = 2)	-.30***	-.01	.36***
Study 3			
Utilitarian PD Parameter	.38***		
Deontology PD Parameter	-.64***	.38***	
Gender (m = 1, f = 2)	-.26**	-.02	.22*

Note: † p = .06, * p < .05, ** p < .01, *** p < .001