

**Sensitivity to Moral Principles Predicts Both Deontological and Utilitarian Response
Tendencies in Sacrificial Dilemmas**

Word count: 4966

D.H. Bostyn¹, A. Roets¹, P. Conway²

¹Ghent University; Department of Developmental, Personality, and Social Psychology; Henri Dunantlaan 2, B-9000, Ghent, Belgium.

²Florida State University, Department of Psychology, 1107 W. Call St., Tallahassee, Florida, USA, 32306-4301. Phone: (850) 644-4079.

Dries.Bostyn@Ugent.be, ORCID: 0000-0001-9994-4615

Arne.Roets@Ugent.be, ORCID: 0000-0001-5814-1189

conway@psy.fsu.edu, ORCID: 0000-0003-4649-6008

Authors' Note

Correspondence concerning this article should be addressed to Dries H. Bostyn, Department of Developmental, Personality, and Social Psychology, Henri Dunantlaan 2, B-9000, Ghent, Belgium. E-mail: Dries.Bostyn@Ugent.be.

Authors' Note: This research was supported by a postdoctoral research grant from the Research Foundation - Flanders (FWO.3E001619) to the first author.

Abstract

When facing sacrificial dilemmas in which harm maximizes outcomes, people appear sensitive to three moral principles: They are more averse to actively causing harm than passively allowing it (*action principle*), more averse to causing harm directly than indirectly (*contact principle*), and more averse to causing harm as a means than as a by-product of helping others (*intention principle*). Across five studies and a meta-analysis ($N = 1218$), we examined whether individual differences in people's sensitivity to these principles were related to participants' moral preferences on sacrificial dilemmas. Interestingly, sensitivity to each of these principles was related to both elevated harm-rejection (i.e., deontological) as well as elevated outcome maximization (i.e., utilitarian) response tendencies. Rather than increasing responses consistent with only one philosophical position, people sensitive to moral principles balanced moral concerns about causing harm and maximizing outcomes, similar to people high in other measures of moral concern.

Keywords: Moral dilemmas; Process dissociation, Deontology; Utilitarianism; Rules; Principles; Action principle; Contact principle; Intention principle

In an oft-quoted passage of his *Critique of Practical Reason*, the German philosopher Immanuel Kant confessed his everlasting admiration for both “the starry heavens above and the moral law within” (p. 1, Guyer, 1992). Kant was touched, not just by the distant glimmer of faraway celestial bodies, but by the strict mathematical laws guiding their movement. Inspired by Newton’s success in reducing the complexity of nature into a set of clearly defined physical principles, Kant aimed to undertake a similar project for morality. However, despite Kant’s best efforts, the principles of normative ethics remain the subject of great philosophical debate. While empirical work cannot address which principles are normatively correct (Ayer, 1936; Gowans, 2016), such research does demonstrate that laypeople are, in fact, sensitive to certain moral principles.

For instance, people judge directly causing harm as worse than passively allowing the same harm to occur, a principle known as the *action principle* (Spranca et al., 1991; Baron & Ritov, 2004; Bostyn & Roets, 2016). Additionally, people think it is more reprehensible to harm someone as a means towards achieving a goal than to inflict the same harm as a side-effect of achieving the same goal, referred to as the *intention principle* (or doctrine of double effect, Foot, 1967; Royzman & Baron, 2002; Waldmann & Dieterich, 2007). Finally, the *contact principle* suggests that harms enacted through direct physical contact are more blameworthy than harms enacted indirectly (Cushman, 2013; Cushman et al., 2006; Greene et al., 2009). While people are, on average, sensitive to these principles, there remains considerable variation across individuals. So why do some people channel their inner Kantian whereas others do not?

One possibility is that sensitivity to these principles is related to the type of moral judgments that participants favor more generally. Considerable research investigates responses to sacrificial dilemmas where harming a focal target minimizes total overall harm (e.g., Greene et

al., 2001; Greene et al., 2004; Kahane, et al., 2015; Conway, et al., 2018).¹ A prime example of these dilemmas, the *trolley dilemma*, describes a runaway trolley that is about to kill five innocent bystanders and asks participants to judge the appropriateness of sacrificing another bystander instead. Typically, harm rejection decisions are described as *deontological*, as they are broadly consistent with philosophical positions advocating absolute prohibitions against causing harm (Kant, 1785/1959). Conversely, harm-acceptance decisions that maximize outcomes are described as *utilitarian*, as they are broadly consistent with philosophical positions advocating a focus on outcomes (e.g., Mill, 1861/1998). Interestingly, whereas principled moral reasoning is foundational to deontological approaches to morality, utilitarian philosophy explicitly rejects the moral importance of such factors. Accordingly, one can wonder whether principled people prioritize a particular response to sacrificial dilemmas, or whether sensitivity to principles elevates multiple moral considerations, suggesting a ‘deep engagement’ with the moral issue instead.

Past work examining moral principledness has largely focused on self-reports of adherence to deontological or utilitarian principles. Such work shows that people who endorse deontological principles tend to reject sacrificial harm (Robinson et al., 2015; Piazza & Landy, 2013), whereas people who endorse utilitarian principles tend to accept sacrificial harm (Conway et al., 2018; Kahane et al., 2018; Piazza & Sousa, 2014). However, effect sizes are small, as people often decide differently than their self-reports may suggest. As a result, this research shows that dilemma decisions reflect a multiplicity of processes rather than an explicit adherence to philosophical ideals. To what extent it is useful to label harm-rejection or outcome-

¹ Some theorists argue for expanding dilemma work to actions that prevent harm to a focal target at a cost to a broader group (e.g., Crone & Laham, 2017; Gawronski et al., 2017). As such focal actions entail a different psychology than harmful actions (Cushman et al., 2012), we focus here on harmful actions that maximize outcomes.

maximizing judgments as ‘deontological’ or ‘utilitarian’ is a matter of active debate (Kahane et al., 2015, 2018; Conway et al., 2018).

In contrast to prior research, we examine principledness by investigating how individual differences in sensitivity to the action, intention, and contact principles relate to sacrificial decisions. Furthermore, we do this by using ‘process dissociation’, a technique allowing for independently assessing harm rejection (i.e. deontological) and outcome-maximization (i.e. utilitarian) response tendencies (Conway & Gawronski, 2013). This procedure allows us to investigate whether people who are more sensitive to these specific principles demonstrate increases in harm-rejection responses tendencies, decreases in outcome-maximization responses tendencies, or increases in both response tendencies.

Principledness and Dilemma Decision-making

Some authors have argued that people sensitive to moral principles, such as the action principle, are making an inherently deontological distinction focused on the harmful action itself (e.g., Baron & Goodwin, 2020; Bennis et al., 2010; cf. Gawronski et al., 2020). Indeed, perhaps the most well-known of deontological principles: “first, do no harm” directly relates to the action principle. This perspective might suggest that people sensitive to moral principles may show increased deontological responding—i.e., rejecting sacrificial harm regardless of what positive outcomes may result. Such findings align with past work suggesting that deontological responding partially reflects adherence to absolute rules, such as inviolable divine commands (Bartels, 2008; Nichols & Mallon, 2006; Piazza & Landy, 2013; Robinson et al., 2015).

A complimentary possibility is that sensitivity to principles is related to reduced utilitarian responding. The three principles are deontological distinctions regarding the nature of

harmful action, and utilitarian philosophy explicitly rejects the moral importance of such factors (Mill, 1861/1998). Therefore, participants inclined towards outcome-focused, consequentialist judgments might demonstrate reduced sensitivity to moral principles. If so, there should be a negative relationship between sensitivity to principles and utilitarian responding.

It is worth pointing out that methodologically, these first two hypotheses are indistinguishable using conventional analyses that treat deontological responding as the direct opposite of utilitarian responding: Both hypotheses suggest that sensitivity to principles would be associated with reduced acceptance of outcome-maximizing harm. However, process dissociation analyses independently assess harm-rejection responding (deontology parameter) and outcome-maximization responding (utilitarian parameter, Conway & Gawronski, 2013). Hence, process dissociation can test both hypotheses simultaneously: whether sensitivity to principles predicts a) increased deontological, and/or b) reduced utilitarian responding.

Moreover, process dissociation allows for testing a third hypothesis: sensitivity to principles might predict both increased deontological and increased utilitarian responding. While such a hypothesis might appear counter-intuitive, consider work that suggests careful deliberative processing increases utilitarian responding (e.g., Byrd & Conway, 2019; Patil et al., 2020; but see also Bostyn et al, 2020). Sensitivity to a moral principle requires the appraisal that that specific principle is relevant to the situation (Nichols & Mallon, 2006). Crucially, the action, intention, and content principles are subtle distinctions some people fail to recognize (Cushman et al., 2006; Baron & Goodwin, 2020). People who take the time to carefully deliberate on moral issues may be more skilled at detecting when these distinctions are relevant, and might also display increased sensitivity to these principles.

Furthermore, a growing body of work demonstrates that people especially concerned about morality score high on both the deontological and utilitarian parameters. For example, people higher in moral identity internalization, which reflects an increased concern for morality at the core of their self-concept (Aquino & Reed, 2002), score high on both parameters, as do people who report high moral conviction against harm (Conway et al., 2018), strong emotional aversion to witnessing others suffer (Reynolds & Conway, 2018), and stable childhoods allowing development of deeper emotional concern for others (Maranges et al., 2020). Hence, people high in moral concern demonstrate deep engagement with and conflict over sacrificial dilemmas, similar to people situationally induced to distrust their initial snap judgment (Conway, Weiss, et al., 2018; Muda et al., 2018; Mata, 2019). Sensitivity to moral principles may operate similarly to these other measures of moral concern, demonstrating associations with both increased deontological and utilitarian responding suggestive of deep engagement with dilemmas.

Note that associative patterns such as these, i.e. scoring high on both deontological and utilitarian parameters, can lead to *suppression* effects. If a variable positively relates to both harm rejection and outcome maximization, the underlying associative patterns will not emerge on a conventional dilemma analysis that treats aversion to harm and maximizing outcomes as opposites because these effects will largely cancel each other out. This is because people who care deeply about both avoiding harm and maximizing outcomes must ultimately arrive at a decision, same as people who care little about either.

Accordingly, if Hypothesis 1, Hypothesis 2, or both are correct, then decision sensitivity to moral principles should emerge for both a conventional dilemma analysis (i.e., fewer sacrificial judgments) and process dissociation analysis (i.e., increased deontology parameter, reduced utilitarian parameter, or both). However, if instead Hypothesis 3 is correct (i.e., people

sensitive to moral principles show both increased deontological and increased utilitarian responding), then this pattern could result in suppression, largely canceling out for conventional analyses that treat harm rejection and outcome maximization as opposites. Process dissociation, however, would still be able to separate and hence detect both effects as increases on both parameters.

The Present Research

We conducted five studies assessing individual differences in decision sensitivity to the *action*, *intention*, and *contact* principles and related these to utilitarian and deontological responding as assessed via process dissociation. Instead of reporting each study individually, we provide the individual results within a comprehensive meta-analysis of all studies. We report all studies collected, all data exclusions and all measures and manipulations used in these studies. The data, materials, detailed results, and statistical code for all studies are available through open science framework (https://osf.io/qxztn/?view_only=180112fb81ec41c2bbb300bdfdd2698b). Studies 1 to 4 were not preregistered, but we preregistered Study 5 at <https://osf.io/h29qj/>.

Method

Participants

For Study 1, we recruited volunteers through social media from the general population of [European country]. For Studies 2-5, we recruited North-American participants through Amazon's Mechanical Turk. We paid participants US\$1.20 for Studies 2 and 4, US\$0.70 for Study 3, and US\$1.50 for Study 5.² We conducted all studies online through inhouse survey software (Studies 1-4) or Qualtrics (Study 5). Participants were barred from completing a study

² Some of these studies were conducted before we were aware of issues involving fair payment of online data-collection and regrettably, this means the payment for some of these studies might have fallen short of meeting the standard of minimum federal wage of U.S. \$7.25.

if they participated previously or duplicated the IP address of a previous participant. We only recruited participants with an HIT approval > 95% but did not require a minimum number of prior HITs. Additionally, for Study 5 we blocked suspicious geocodes, and used CloudResearch to verify worker location and block low quality participants. No participants were excluded from Studies 1-4. Study 5 contained an attention check that required checking a specific multiple choice answer and copying “I read the instructions” to a text box. Participants that failed this check ($n = 22$) were eliminated from the sample.

Basic demographic statistics and sample sizes for each study are available in Table 1. One participant in Study 1 self-reported an age of 2 which was recoded to missing. We did not collect any data on the racial composition of our samples in Studies 1 to 4. In Study 5, 75% of participants identified as White, 9% identified as Asian, 7% identified as Black, 6% identified as Hispanic, 2% identified as Multi-racial and <1% of participants identified as American Indian or Alaskan Native.

We conducted an internal meta-analysis of all five studies, but not all studies contained all measures (See table 2). A power sensitivity analysis suggests the meta-analytic test associated with the smallest sample (i.e. the correlation between sensitivity to principles and process dissociation parameters; $N = 1076$) had ~80% power to detect an effect of $r \geq .11$, assuming moderate between-study heterogeneity (Borenstein et al., 2011; Valentine et al., 2010).

Table 1

Sample size and summary demographical statistics

	<i>N</i> (Female, Male, Other)	Mean Age (<i>SD</i>)
Study 1	120 (76, 44, 0)	33.52 (15.6)
Study 2	199 (96, 103, 0)	35.48 (9.6)
Study 3	200 (96, 104, 0)	35.27 (10.9)
Study 4	197 (87, 110, 0)	35.82 (11.6)
Study 5	480 (216, 262, 2)	39.47 (11.8)

Procedure

We used a similar procedure throughout all studies, albeit with minor differences. In all studies, participants were first asked to complete demographic information (age and gender, plus racial background in Study 5). In Study 1, we then measured decision sensitivity to the three moral principles on a randomized battery of eight sacrificial dilemmas. Each dilemma entailed asking participants how morally appropriate it was to sacrifice a single bystander to save five others, but differed on whether this choice entailed completing an action or refraining from an action (i.e. action principle), whether sacrificing the bystander served as a means towards saving the others or was a side-effect of saving the others (i.e. intention principle), and finally, whether direct or indirect harmful contact was involved (i.e. contact principle). Then, participants completed a randomized battery of six different sacrificial dilemmas where causing harm maximizes outcomes (Bostyn et al., 2016). This battery allows for a conventional dilemma analysis. Finally, participants reported whether they explicitly endorsed the three moral principles.

In Study 2, participants first completed a randomized 20-item sacrificial dilemma battery, with 10 congruent and incongruent variants, allowing for both a conventional and process dissociation analysis (Conway & Gawronski, 2013). Participants then completed the 8 dilemmas from Study 1 to measure decision sensitivity to moral principles. Finally, participants completed two additional trait measures: Need for Cognition (Cacioppo & Petty, 1982; Cacioppo et al., 1984) and Need to Belong (Baumeister & Leary, 1995). Study 2 did not include a measure of explicit agreement with principles.

In Studies 3, 4 and 5 participants again completed the process dissociation battery and moral principle decision sensitivity battery (cf. Study 2), and participants also reported explicit

endorsement of each moral principle. In Studies 3 and 4, participants also completed the Rational-Experiential Inventory (Epstein et al., 1996). The REI combines a measure of Need for Cognition with a measure for Experiential Thinking. Study 5 did not contain any trait measures but did additionally ask participants whether they had encountered trolley-like dilemmas before.³ Full results for the trait measures and the measure of explicit approval are available in the supplementary materials.⁴ An overview of which measures were included in which study is available in Table 2.

Table 2

Measures included in each experiment.

	Study 1	Study 2	Study 3	Study 4	Study 5
Traditional Dilemmas	✓				
Process Dissociation Dilemmas		✓	✓	✓	✓
Sensitivity to Principles	✓	✓	✓	✓	✓
Additional Measures:					
Explicit Approval	✓		✓	✓	✓
Need for Cognition		✓			
Need to Belong		✓			
REI			✓	✓	

Focal Measures

³ Results are similar for both groups, a more complete analysis is available in the supplementary materials.

⁴ Neither Need to Belong nor Experiential thinking were associated with usage of any moral distinction (all $|r| < .07$, $p > .095$). Need for Cognition was associated with increased sensitivity to the action principle $r_{meta} = .08$, $p = .033$, increased sensitivity to the intention principle, $r_{meta} = .12$, $p = .02$ but not with sensitivity to the contact principle, $r_{meta} = -.01$, $p = .401$. Explicit approval of a moral principle correlated with sensitivity to that principle in the case of the action principle, $r_{meta} = .11$, $p = .018$, and the contact principle, $r_{meta} = .11$, $p = .004$, but not in the case of the intention principle, $r_{meta} = .03$, $p = .201$.

Decision sensitivity to moral principles. Participants responded to eight variants of the trolley dilemma. Each dilemma involved the tradeoff of sacrificing one life to save five lives, and participants indicated to what extent they deemed this sacrifice to be morally appropriate on 7-point scales from (1) *completely inappropriate* to (7) *completely appropriate* in Study 1, and 5-point scales from (1) *completely inappropriate* to (5) *completely appropriate* in Studies 2-5.⁵ Using a 2×2×2 within-subjects design, each participant considered eight dilemmas comprising all possible combinations of the three distinctions in a randomized fashion. In four dilemmas sacrificing the single bystander required an action; in the other four, it required inaction. In four dilemmas the sacrifice required using the single person as the means of saving the five; in the other four, the sacrifice was only a side-effect. Finally, in four dilemmas the sacrifice necessitated direct physical contact; in the other four, it did not. As an example: the “Action/Means/Contact” dilemma read as follows:

“A runaway trolley-train is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge stands another workman. The only way to save the lives of all five workmen is by pushing this other workman off the bridge onto the tracks where his body will sufficiently slow down the trolley so all five workmen can be safely evacuated. If you do this, the one workman will die but the five others will be saved. How morally appropriate is it to sacrifice the one workman to save the five?”

Whereas, the “Inaction/Means/Contact” dilemma read:

⁵ In Study 1, participants also rated how appropriate they considered it to *not* sacrifice the single bystander (as per Bostyn et al., 2018). Because the results from this measure mirrored those reported in the main text, we simplified our procedure for all subsequent studies. These results are available in the online supplementary materials.

“A single workman is stuck between the rails of a railroad track. You grab his hand and are about to free the man when you notice a runaway trolley-train headed towards the stuck workman, and some distance away, five other workmen. The only way to save the lives of the other five workmen is to not free the single workman so his body will sufficiently slow down the trolley so all five workmen can be safely evacuated. If you do this, the one workman will die but the five others will be saved. How morally appropriate is it to sacrifice the one workman to save the five?”

Conventional and process dissociation dilemma batteries. In the first study, participants considered six sacrificial moral dilemmas (Bostyn et al., 2016) and indicated whether they considered the sacrificial harm to be either morally appropriate or inappropriate in a binary fashion (*yes/no*). This battery allows for a conventional analysis only. In Studies 2 to 5, we employed process dissociation. Process dissociation also allows for a conventional analysis directly comparable to Study 1, so we ran a) the conventional analysis for all five studies, and b) a process dissociation analysis on Studies 2-5.

Computing process dissociation parameters requires participants to respond to two types of sacrificial dilemmas: *congruent* and *incongruent*. Incongruent dilemmas entail causing harm to maximize overall outcomes (e.g. killing a baby to save the lives of many villagers). Similar to traditional trolley-style dilemmas, concerns about avoiding harm and maximizing outcomes are incongruent. Congruent dilemmas involve the same scenario and sacrificial harm, but the positive benefits from inflicting harm are diluted so they are harder to justify on utilitarian grounds (killing a baby to save the same villagers from hard labor). For each harmful action with a specified outcome, participants selected either *Yes, this is appropriate*, or *No, this is not appropriate*. By computing acceptance and rejection of sacrificial harm for each dilemma for

each participant, and applying the formulae described by Conway and Gawronski (2013), we calculated the degree to which each participant rejected sacrificial harm across all scenarios regardless of outcomes (deontology parameter), and the degree to which they accepted harm when doing so maximizes outcomes whilst rejecting harm when it does not (utilitarian parameter). A full explanation of process dissociation calculations is available through the online supplementary materials.

Results

Decision Sensitivity to Moral Principles

First, we measured whether decisions on the 8-dilemma principle battery varied in line with each principle—i.e., whether participants' judgments demonstrated the action, intention, and contact principles. To compute the action principle sensitivity score, we averaged sacrificial acceptance ratings for inaction dilemmas and subtracted average acceptance ratings from action dilemmas. Higher scores denote increased sensitivity to the principle. We used a similar procedure to compute sensitivity to the intention and contact principles. Overall, participants showed significant sensitivity to all three moral principles. Across all studies, they were more likely to consider sacrificial harm inappropriate when harm resulted from actions, served as a means towards a goal, or was accomplished through physical contact, consistent with past work (see Table 2; Cushman et al., 2006; Feltz & May, 2017; Jamison et al., 2020).

Table 3

Mean and standard deviations of decision sensitivity to each moral principle for all studies. Sensitivity to principles was computed as a difference score based on measures with a 7-point scale in Study 1 and a 5-point scale in Studies 2–5.

Action Principle				
	Mean (SD)	<i>t</i>	<i>p</i>	Cohen's <i>D</i> (95% CI)
Study 1 (<i>N</i> = 120)	0.67 (0.90)	8.11	<.001	0.74 (0.58, 0.92)
Study 2 (<i>N</i> = 199)	0.30 (0.60)	7.05	<.001	0.50 (0.38, 0.62)
Study 3 (<i>N</i> = 200)	0.24 (0.56)	5.98	<.001	0.42 (0.30, 0.56)
Study 4 (<i>N</i> = 197)	0.27 (0.66)	5.70	<.001	0.41 (0.29, 0.54)
Study 5 (<i>N</i> = 480)	0.18 (0.54)	7.42	<.001	0.34 (0.25, 0.43)
Meta:				0.43 (0.27, 0.59)
Intention Principle				
Study 1 (<i>N</i> = 120)	0.58 (0.76)	8.45	<.001	0.77 (0.60, 0.98)
Study 2 (<i>N</i> = 199)	0.18 (0.52)	4.91	<.001	0.35 (0.22, 0.48)
Study 3 (<i>N</i> = 200)	0.13 (0.48)	3.73	<.001	0.26 (0.13, 0.39)
Study 4 (<i>N</i> = 197)	0.18 (0.56)	4.60	<.001	0.33 (0.20, 0.44)
Study 5 (<i>N</i> = 480)	0.13 (0.50)	5.86	<.001	0.27 (0.18, 0.36)
Meta:				0.34 (0.14, 0.54)
Contact Principle				
Study 1 (<i>N</i> = 120)	0.96 (1.06)	9.95	<.001	0.91 (0.75, 1.08)
Study 2 (<i>N</i> = 199)	0.30 (0.57)	7.46	<.001	0.53 (0.42, 0.64)
Study 3 (<i>N</i> = 200)	0.32 (0.58)	7.91	<.001	0.56 (0.45, 0.68)
Study 4 (<i>N</i> = 197)	0.28 (0.56)	7.07	<.001	0.50 (0.37, 0.64)
Study 5 (<i>N</i> = 480)	0.22 (0.58)	8.48	<.001	0.39 (0.30, 0.48)
Meta:				0.51 (0.30, 0.72)

Conventional Dilemma Analyses

Next, to examine whether individual differences in sensitivity to each principle predicted responses on sacrificial dilemmas, we calculated the correlations between these measures in each study and combined them in a meta-analysis with the *metacor* package in R (Laliberté, 2011) using a DerSimonian-Laird approach (see Figure 1). For conventional analyses that treat deontological and utilitarian concerns as opposites, we uncovered some evidence that increased sensitivity to the action principle was related to reduced utilitarian dilemma responses, $r_{meta} = -.07$, $p = .012$, but no evidence for an association between sensitivity to the intention or contact principles and utilitarian dilemma responses, both $|r_{meta}| \leq 0.01$, both $ps > .475$. However, as discussed above, such analyses are vulnerable to suppression effects: sensitivity to moral principles may predict increased deontological *and* increased utilitarian responding. If so, then we should expect suppression to result in low correlations with conventional analyses, similar to other studies (e.g., Maranges et al., 2020).

Process Dissociation Analyses

We calculated the correlation between sensitivity to each principle with the deontology and utilitarian PD parameters for Studies 2-5, and combined these results into a meta-analysis (see Figure 1). Consistent with Hypothesis 3, results showed that for each moral principle, sensitivity to that principle significantly predicted both higher deontological responding, all $r_{meta} > .12$, $p \leq .001$, and higher utilitarian responding, all $r_{meta} > .15$, $p \leq .001$. People sensitive to the action, intention, and/or contact principles tended to both reject sacrificial harm and maximize overall outcomes, once these response tendencies were disentangled. Although results were not

always significant within each individual study, overall, the pattern was similar across all principles and studies. Notably, this pattern replicated in preregistered Study 5.

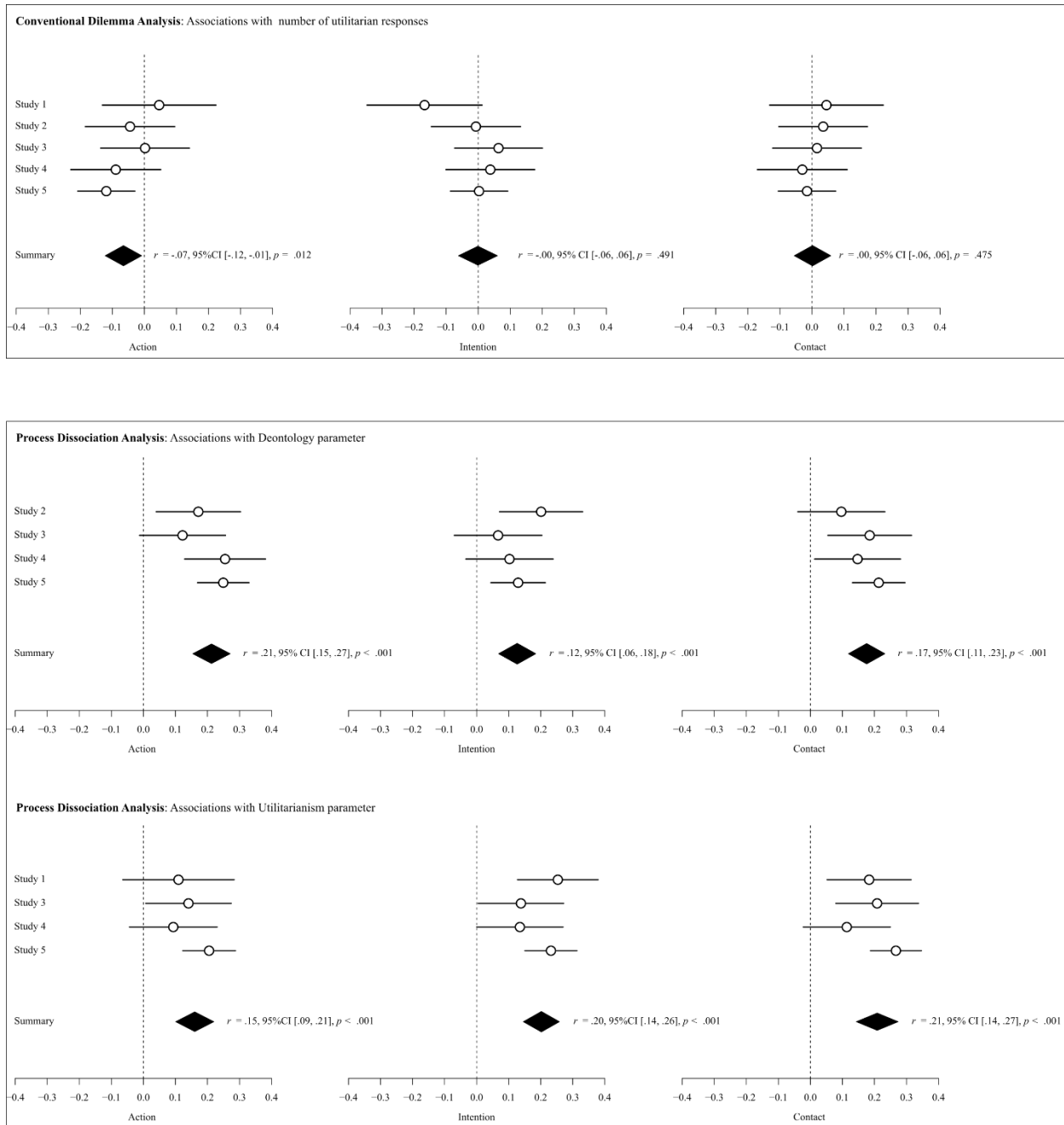


Figure 1. Meta-analytic correlations across all five studies for sensitivity to principles and moral judgments (upper panel: conventional analysis, lower panel: process dissociation analysis).

General Discussion

We investigated how individual differences in sensitivity to the action, intention, and contact principles relate to responses on sacrificial dilemmas. Overall, participants' decisions were clearly sensitive to each principle. These results thereby constitute a robust replication of earlier research in this area (Baron & Ritov, 2004; Cushman et al, 2006; Cheng, 2020; Feltz & May, 2017; Jamison et al., 2020). Yet, not everyone was equally sensitive to each principle. We then examined how individual differences in sensitivity to these principles relate to responses on subsequent sacrificial dilemmas using conventional (Studies 1-5) and process dissociation analyses (Studies 2-5).

Conventional analyses revealed only a single small effect: When concerns about rejecting harm were pitted against concerns about maximizing outcomes, a negative correlation emerged between sacrificial acceptance and sensitivity to the action principle (only). Conclusions based on this conventional analysis might therefore suggest that sensitivity to moral principles is largely unrelated to increased deontological responding (Hypothesis 1), reduced utilitarian responding (Hypothesis 2), or increased utilitarian and deontological responding (Hypothesis 3). However, conventional analyses can be misleading in cases of suppression. By independently assessing the degree to which each participant rejects sacrificial harm regardless of outcomes (deontological parameter), and accepts harm that maximizes outcomes (utilitarian parameter), process dissociation (Conway & Gawronski, 2013) allows for a more fine-grained analysis.

Importantly, the process dissociation analysis supported Hypothesis 3: participants more sensitive to the action, intention, or contact principle scored higher on both the deontology and utilitarian parameters. In other words, participants that paid careful attention to morally relevant factors pertaining to action, intention, and contact, strove to balance harm rejection and outcome-

maximization. This pattern also explains the results for conventional measures: both positive associations counteracted one another, either washing out the overall effect (i.e., suppression), or resulting in a small negative one.

Principled deontologists and principled utilitarians

To our knowledge, the current work is the first to explore individual differences in sensitivity to the action, intention, and contact principles. Participants sensitive to each principle demonstrated both a heightened aversion to causing harm and a heightened desire to maximize outcomes. These results suggest that people sensitive to principles engage deeply with dilemmas, carefully considering both the well-being of the sacrificial target and the group. Consistent with this interpretation, past work demonstrated similar heightened utilitarian and deontological responding among people high in markers of moral concern such as moral identity internalization (Conway et al., 2018; Reynolds & Conway, 2018), and when context promoted close and careful reading of dilemmas (Conway, Weiss, et al., 2018; Muda, et al., 2018).

Importantly, these findings are inconsistent with theories suggesting that principledness uniquely increases deontological responding or reduces utilitarian responding. People sensitive to moral principles appear to balance multiple moral concerns rather than prioritizing only one. Research on sacrificial harm has long used philosophical labels to describe harm aversive and outcome-maximizing judgments. While such labels can be defended on pragmatic and scientific grounds (Conway et al, 2018), these labels might also tempt people to assume that deontological judgments may be driven primarily by moral principles, whereas utilitarian judgments may be determined by outcomes alone. Our data demonstrates that this assumption is inexact: participants who favor utilitarian judgments display equally strong tendencies towards principled

moral judgments (as measured by their usage of the action, intention, and contact principle) as those who favor deontological judgments do.

There might be practical implications to these findings as well. During the COVID-19 pandemic, problems of sacrificial harm became a matter of public debate. In the initial months of the global vaccination drive, evidence arose that certain vaccines could trigger a rare blood clotting issue. Although preliminary evidence suggested that these side effects were rare and less common than comparable side-effects from COVID-19 infection (Taquet et al., 2021), several countries nonetheless suspended useage of these vaccines. As these decisions were anti-utilitarian, they were widely critiqued. However, harms from COVID-19 are harms of omission, whereas harms by vaccination are harms of commission. This complicates the relative comparison of these harms. The current work suggests that people sensitive to the omission/commission distinction may not devalue utilitarian considerations and may even experience heightened utilitarian concerns, which they struggle to balance against heightened deontological concerns. Accordingly, even when utilitarian calculus clearly suggests one possible policy response, it remains important for public health officials to weigh other moral concerns as well. Proposed solutions may find broader public support if they explicate concern for both utilitarian and deontological considerations (see Rom et al., 2017).

Limitations

Like all work, the current studies have limitations. First, the current work was conducted in Europe and North-America, thereby only including ‘WEIRD’ participants (Henrich et al., 2010). There are cross-cultural differences in the extent to which people embrace certain moral principles such as the action principle (Awad et al., 2020). Accordingly, we caution against

extrapolating the current results beyond their specific cultural context. Second, while the demonstrated associations are theoretically meaningful, the effects we have uncovered are small to medium-sized. Sensitivity to these principles appears to only be modestly related to preferences for deontological or utilitarian resolutions. Third, like most sacrificial dilemma research, we employed dilemmas that are hypothetical (Bauman et al., 2014), not always plausible (Körner, et al., 2019), and with questionable real-world predictive validity (Bostyn, et al., 2018). However, such dilemmas do remain useful for clarifying the cognitive structure of moral psychology (Cushman & Greene, 2012).

Finally, like most sacrificial dilemma research, we examined dilemmas where utilitarian decisions necessitate action and deontological decisions necessitate inaction. Some authors have argued that this action/inaction distinction confounds concern for a focal target with a general unwillingness to act (Gawronski & Beer, 2017; Gawronski et al., 2017). Accordingly, these authors argue in favor of explicitly incorporating action/inaction tendencies as a third independent moral concern at the same level as response tendencies toward harm aversive or outcome-maximizing judgment. It remains unclear whether the current findings would be reproduced using such a model. Importantly, recent work using this model largely corroborates our argument that moral concern entails balancing multiple considerations: for example, people high in moral identity score high on all three parameters, demonstrating not only concern for harm aversion and outcomes, but also inaction (Körner et al., 2020). Hence, an exciting question for future work would be whether people sensitive to moral principles balance multiple moral considerations in these models. However, it is important to note that these models are contested as well. Other researchers have argued that the action/inaction distinction intrinsically typifies deontological judgment, where action that harms others are strictly forbidden and actions that aid

others are merely laudatory (Baron & Goodwin, 2020) and thus, that it is theoretically flawed to envision action/inaction tendencies as a separate dimension. This debate is not yet settled and readers are cautioned to take this issue into advisement when interpreting our results.

Conclusion

The current research demonstrates that individual differences in sensitivity to moral principles are associated with a balanced approach to concerns for individual victims and the larger group in sacrificial moral dilemmas. Hence, such individual differences could help refine insight into moral cognition. Whereas most moral psychology research has focused on either studying participants' overall dilemma decision-making, or on subtle effects of each moral principle, these data demonstrate the added value of combining both perspectives.

References

- Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J. F. (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proceedings of the National Academy of Sciences*, 117(5), 2332-2337.
- Ayer, A. J. (1936). *Language, truth, and logic*. Gollancz.
- Baron, J., Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational behavior and human decision processes*, 94, 74–85.
- Baron, J., & Goodwin, G. P. (2020). Consequences, norms, and inaction: A critical analysis. *Judgment & Decision Making*, 15(3).
- Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, 108, 381-417.
- Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536-554.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117, 497-529.
- Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, 5, 187–202.
doi:10.1177/1745691610362354
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

- Bostyn, D. H., & Roets, A. (2016). The morality of action: The asymmetry between judgments of praise and blame in the action–omission effect. *Journal of Experimental Social Psychology, 63*, 19-25.
- Bostyn, D. H., Roets, A., & Van Hiel, A. (2016). Right-wing attitudes and moral cognition: Are Right-Wing Authoritarianism and Social Dominance Orientation related to utilitarian judgment?. *Personality and individual differences, 96*, 164-171.
- Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological science, 29*, 1084-1093.
- Bostyn, D. H., De Keersmaecker, J., Van Assche, J., & Roets, A. (2020). Bright mind, moral mind? Intelligence is unrelated to consequentialist moral judgment in sacrificial moral dilemmas. *Psychonomic bulletin & review, 27*, 392-397.
- Byrd, N., & Conway, P. (2019). Not all who ponder count costs: arithmetic reflection predicts utilitarian tendencies, but more general reflection predicts both deontological and utilitarian tendencies. *Cognition, 195*, 1-19.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology, 42*, 116-131.
- Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of personality assessment, 48*(3), 306-307.
- 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.

- Conway, P., Goldstein-Greenwood, J., Polacek, D., and Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition*, *179*, 241-265.
- 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.
- Crone, D. L., & Laham, S. M. (2017). Utilitarian preferences or action preferences? Deconfounding action and moral code in sacrificial dilemmas. *Personality and Individual Differences*, *104*, 476-481.
- Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and social psychology review*, *17*, 273-292.
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*, 2-7.
- Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social neuroscience*, *7*(3), 269-279.
- 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*(11), 1082-1089.
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of personality and social psychology*, *71*, 390-405.
- Feltz, A., & May, J. (2017). The means/side-effect distinction in moral cognition: A meta-analysis. *Cognition*, *166*, 314-327.

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

Gawronski, B., & Beer, J. S. (2017). What makes moral dilemma judgments “utilitarian” or “deontological”? *Social Neuroscience*, 12, 626-632.

Gawronski, B., Conway, P., Hütter, M., Luke, Dillon, M., Armstrong, J., & Friesdorf, R., & (2020). On the Validity of the CNI Model of Moral Decision-Making: Reply to Baron and Goodwin (2020). *Manuscript under review*.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology*, 113, 343-376.

Gowans, C. (2016). Moral relativism. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/win2016/entries/moral-relativism/>.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111, 364-371.

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(2), 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(5537), 2105-2108.

Guyer, P. (1992). Introduction. In P. Guyer (Ed.), *The Cambridge companion to Kant*. Cambridge University Press.

- Hennig, M., & Hütter, M. (2020). Revisiting the divide between deontology and utilitarianism in moral dilemma judgment: A multinomial modeling approach. *Journal of personality and social psychology, 118*, 22-56.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences, 33*(2-3), 61-83.
- Jamison, J., Yay, T., & Feldman, G. (2020). Action-inaction asymmetries in moral scenarios: Replication of the omission bias examining morality and blame with extensions linking to causality, intent, and regret. *Journal of Experimental Social Psychology, 89*, 103977.
- Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193-209.
- Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review, 125*, 131-164.
- Kant, I. (1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Bobbs-Merrill. (Original work published 1785).
- Körner, A., Joffe, S., & Deutsch, R. (2019). When skeptical, stick with the norm: Low dilemma plausibility increases deontological moral judgments. *Journal of Experimental Social Psychology, 84*, 103834.
- Körner, A., Deutsch, R., & Gawronski, B. (2020). Using the CNI model to investigate individual differences in moral dilemma judgments. *Personality and Social Psychology Bulletin, 0146167220907203*.

Laliberté, E. (2011). Metacor: Meta-analysis of correlation coefficients. R package version 1.0-2.

Retrieved from <https://CRAN.R-project.org/package=metacor>

Maranges, H. M., Hasty, C. R., Maner, J. K., & Conway, P. (2020). The behavioral ecology of moral dilemmas: Childhood unpredictability, but not harshness, predicts less deontological and utilitarian responding. *Journal of Personality and Social Psychology*. *In Press*.

Mata, A. (2019). Social metacognition in moral judgment: Decisional conflict promotes perspective taking. *Journal of Personality and Social Psychology*. 1061-1082.

Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). New York, NY: Oxford University Press. (Original work published 1861).

Muda, R., Niszczoła, P., Białek, M., & Conway, P. (2018). Reading dilemmas in a foreign language reduces both deontological and utilitarian response tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2), 321.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100(3), 530-542.

Patil, I., Zucchelli, M. M., Maria, M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2020). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology*. *In Press*.

Piazza, J., & Landy, J. F. (2013). "Lean not on your own understanding": Belief that morality is founded on divine authority and non-utilitarian moral judgments. *Judgment and Decision-making*, 8, 639-661.

Piazza, J., & Sousa, P. (2014). Religiosity, political orientation, and consequentialist moral thinking. *Social Psychological and Personality Science*, 5, 334-342.

- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reynolds, C. J., & Conway, P. (2018). Not Just Bad Actions: Affective Concern for Bad Outcomes Contributes to Moral Condemnation of Harm in Moral Dilemmas. *Emotion*, 18(7), 1009-1023. doi: 10.1037/emo0000413
- Robinson, J. S., Joel, S., & Plaks, J. E. (2015). Empathy for the group versus indifference toward the victim: Effects of anxious and avoidant attachment on moral judgment. *Journal of Experimental Social Psychology*, 56, 139-152.
- Royzman, E., Baron, J. (2002). The preference for indirect harm. *Social justice research*, 15, 165–184.
- Spranca, M., Minsk, E., Baron, J. (1991). Omission and commission in judgment and choice. *Journal of experimental social psychology*, 27, 76–105.
- Taquet, M., Husain, M., Geddes, J. R., Luciano, S., & Harrison, P. J. (2021). Cerebral venous thrombosis and portal vein thrombosis: a retrospective cohort study of 537,913 COVID-19 cases. *medRxiv*. doi: 10.1101/2021.04.27.21256153.
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(2), 215-247.
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological science*, 18(3), 247-253.