



The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

Judgments of moral responsibility and wrongness for intentional and accidental harm and purity violations

Mary Parkinson & Ruth M. J. Byrne

To cite this article: Mary Parkinson & Ruth M. J. Byrne (2017): Judgments of moral responsibility and wrongness for intentional and accidental harm and purity violations, The Quarterly Journal of Experimental Psychology, DOI: 10.1080/17470218.2016.1276942

To link to this article: http://dx.doi.org/10.1080/17470218.2016.1276942

Accepted author version posted online: 06 Jan 2017. Published online: 31 Jan 2017.



🧭 Submit your article to this journal 🗗





View related articles 🗹



則 🛛 View Crossmark data 🗹

Full Terms & Conditions of access and use can be found at http://www.tandfonline.com/action/journalInformation?journalCode=pqje20

Judgments of moral responsibility and wrongness for intentional and accidental harm and purity violations

Mary Parkinson and Ruth M. J. Byrne

School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Ireland

ABSTRACT

Two experiments examine whether people reason differently about intentional and accidental violations in the moral domains of harm and purity, by examining moral responsibility and wrongness judgments for violations that affect others or the self. The first experiment shows that intentional violations are judged to be worse than accidental ones, regardless of whether they are harm or purity violations—for example, Sam poisons his colleague versus Sam eats his dog, when participants judge how morally responsible was Sam for what he did, or how morally wrong was what Sam did. The second experiment shows that violations of others are judged to be worse than violations of the self, regardless of whether they are harm or purity violations, when their content and context is matched—for example, on a tropical holiday Sam orders poisonous starfruit for dinner for his friend, or for himself, versus on a tropical holiday Sam orders dog meat for dinner for his friend, or for himself. Moral reasoning is influenced by whether the violation was intentional or accidental, and whether its target was the self or another person, rather than by the moral domain, such as harm or purity.

Most people probably sympathize with Barack Obama's decision not to drink urine when he appeared on Bear Grylls' TV survival show, and his assertion, "I suppose, in extremis, it's something that I would do-if the alternative was death. It's not something I'd make a habit of. And I probably wouldn't do it just for a TV show" (Carroll, 2015). Many people are disgusted by violations of purity, such as drinking urine, yet they can nonetheless reason about such emotive matters (e.g., Gubbins & Byrne, 2014). People can reason about a wide variety of deontic matters about what they are obliged to do, what they are permitted to do, and what they are forbidden from doing, and an important sort of deontic inference concerns moral matters (e.g., Bucciarelli, Khemlani, & Johnson-Laird, 2008). The two experiments we report aim to examine the cognitive processes that underpin reasoning about intentional and accidental violations in the moral domains of harm and purity; they examine whether such differences arise because of

ARTICLE HISTORY

Received 29 January 2015 Accepted 13 December 2016

KEYWORDS

Harm; Intentions; Moral reasoning; Purity; Responsibility

differences in the way that people think about violations of others compared to violations of the self, and differences in the way they think about the moral responsibility of an agent compared to the moral wrongness of an agent's actions.

Intentional and accidental violations of harm and purity

We examine reasoning about violations in the domains of harm and purity. Moral violations have been categorized within different moral domains such as harm and purity, as well as other domains such as fairness, loyalty, and authority (e.g., Haidt, 2001; Shweder, Much, Mahapatra, & Park, 1997). The "harm" domain is related to the prevention of physical harm and the provision of care and nurturance, and violations often concern hurting other people—for example, a man who poisons his work colleague. The "purity" domain is related to the preservation of

CONTACT Mary Parkinson mary.parkinson@dcu.ie Trinity College Dublin, University of Dublin, Dublin, Ireland
Supplemental data for this article can be accessed here https://reasoningandimagination.wordpress.com/ and at http://www.surveygizmo. com/.

the sanctity of the body, and violations often concern ingesting or otherwise taking into the body potentially contaminating substances—for example, a woman who drinks her flatmate's urine.

We also examine reasoning about violations that are intentional or accidental. In general, moral opprobrium is greatly reduced for agents acting without bad intent (e.g., Cushman, 2008; Gray & Schein, 2012). People readily distinguish between actions that are intentional and actions that are not (e.g., Anscombe, 1963; Bratman, 1987; Davidson, 1963; Von Wright, 1983). They rely on intentions to inform their understanding of an agent's actions not only in moral situations (e.g., Cushman & Young, 2011; Knobe, 2010; Nadelhoffer, 2004; Sverdlik, 2004), but also in nonmoral ones (e.g., Byrne, 2005; Juhos, Quelhas, & Byrne, 2015; Rasga, Quelhas, & Byrne, 2016; Uttich & Lombrozo, 2010).

Our focus on intentional and accidental harm and purity violations arises in part because of recent findings that judgments about harm take into account the intention of the actor more so than judgments about purity (e.g., Young & Saxe, 2011). For example, participants read about a purity violation:

Imagine that a car just killed your beloved dog. Your significant other has heard that dog meat is delicious and freezes the meat of your dog before it goes bad. Later, you decide to make yourself dinner. You see a package in the freezer. It is incorrectly labeled "beef"—but it is actually the meat from your dog. Without realizing it, you end up eating your dog for dinner.

Alternatively, they read about a harm violation:

Imagine you and a co-worker are taking a tour of a chemical plant. During the coffee break, you go to pour some coffee. You like yours black, but your co-worker asks for sugar in hers. You spoon some powder into your coworker's coffee. You do not know that someone replaced the sugar with poison.

The moral violations were described as accidental —the agent did not know what he or she was doing, or they were described as intentional—the agent did know what he or she was doing. Unsurprisingly, when participants were asked, "how morally wrong was the action?" they judged the intentional violation of a moral principle to be very wrong (about 6–7 on a 1–7-point scale where 7 is very morally wrong), and they did so equally for harm and purity violations (Young & Saxe, 2011). However, a curious finding is that participants judged the accidental purity violation to be much more wrong than the accidental harm violation—about 5–6 for the purity violation on the 7-point scale compared to about 1–2 for the harm violation (Young & Saxe, 2011). The two experiments we report examine such differences in the way people reason about intentional and accidental violations in the moral domains of harm and purity.

Self versus other violations, and moral responsibility versus wrongness judgments

We examine whether people think differently about intentional and accidental harm and purity violations because of differences in the way they think about violations that affect others and violations that affect oneself. The target of moral violations typically differs for the different domains of harm and purity; an agent who commits a harm violation typically has an impact on another person-for example, Sam poisons his colleague—but an agent who commits a purity violation typically affects his or her own purity rather than another person's-for example, Sam eats his dog (e.g., Chakroff, Dungan, & Young, 2013; Chakroff & Young, 2015b; Young & Tsoi, 2013; also see Rottman, Kelemen, & Young, 2014). Differences in the target of the violation may contribute to the difference in judgments of accidental purity and accidental harm (e.g., Chakroff et al., 2013). We examine harm and purity violations of others and of the self in situations that are carefully matched for content and context—for example, on a tropical holiday Sam accidentally orders dog meat for himself, or for his friend (a purity violation), or he accidentally orders poisonous starfruit for himself, or for his friend (a harm violation).

We also examine whether people think differently about intentional and accidental harm and purity violations because of differences in the way they think about the moral responsibility of an agent and the moral wrongness of an agent's actions. One suggested explanation of the difference in judgments of accidental purity and accidental harm has been that people might not consider the agent's mental state as much when they think about purity violations as they do when they think about purity violations (e.g., Chakroff et al., 2015; Chakroff & Young, 2015a; Young & Tsoi, 2013). Instead they may focus on the unacceptability of the act itself, and possibly the agent's moral character (e.g., Chakroff et al., 2013; Chakroff & Young, 2015b; Graham, Haidt, & Nosek, 2009). We test this explanation by examining judgments that may direct attention to an agent's mental state, such as "How morally responsible is Sam for what he did?".

Judgments of the responsibility of an agent and judgments of the wrongness of an act are influenced by assessments of the intentions of the agent (e.g., Cushman, 2008). Judgments of moral wrongness often focus on assessments of the moral principle violated, such as the importance of the principle, and the severity of the outcome (e.g., Cushman, 2008; Miller & Cushman, 2013; Miller, Hannikainen, & Cushman, 2014). For example, people often make reference to moral principles when stating that something is wrong-for example, "war is wrong because it conflicts with Christian principles" (e.g., Watson, 1999, p. 64; see Malle, Guglielmo, & Monroe, 2014). Judgments of moral responsibility often focus on assessments of the agent's specific contribution to the violation, such as their causal role, and their social obligations and capacity-for example, what they should have done, or could have done, differently (e.g., Heider, 1958; Lagnado, Gerstenberg, & Zultan, 2013; Schlenker, Britt, Pennington, Murphy, & Doherty, 1994; Shaver, 1985; Shoemaker, 2007; Shultz & Schleifer, 1983; Shultz & Wright, 1985; Weiner, 1995). Moral responsibility has been considered to be the accountability or answerability of an agent for wrongdoing or their eligibility for blame (e.g., Fincham & Jaspers, 1980; Shultz, Wright, & Schleifer, 1986; Shaver, 1985; Weiner, 1995). The two experiments we report examine judgments about violations that affect others and violations that affect oneself, and judgments of the responsibility of the agent and the wrongness of the act, to test how people think differently about intentional and accidental harm and purity violations.

Experiment 1

The aim of the first experiment was to examine reasoning about intentional and accidental harm and purity violations, in judgments about the moral responsibility of the agent—for example, "How morally responsible is Sam for what he did?"—and the moral wrongness of the agent's actions—for example, "How morally wrong was what Sam did?". We relied on four stories about accidental or intentional violations of harm or purity (adapted with minor changes from Young & Saxe, 2011). An example of one of the purity violation stories is the following:

Imagine that a car just killed Sam's dog, mangling her badly. Sam's partner collects the remains and freezes them. Later that week, Sam decides to make himself dinner. He sees several packages in the freezer.

In one version the story resulted in an accidental purity violation:

They are incorrectly labelled "beef" and it is the meat from his dog. Without realizing it, Sam ends up eating his dog for dinner.

In another version it ended in an intentional purity violation:

They are correctly labelled "dog" and it is the meat from his dog. Knowingly, Sam ends up eating his dog for dinner.

The other purity violation story was about drinking urine (see the Appendix). An example of one of the harm violation stories is the following:

Imagine Joe and a co-worker are taking a tour of a chemical plant. During the coffee break, Joe goes to pour some coffee. Joe likes his black, but his co-worker asks for sugar in hers.

In one version the story ended in an accidental harm violation:

He does not know that he is putting poison in her coffee. Joe spoons some powder into his co-worker's coffee.

In the other version it ended in an intentional harm violation:

He knows that he is putting poison in her coffee. Joe spoons some powder into his co-worker's coffee.

The other harm violation story was about giving an allergen to a family member (see the Appendix).

Method

Participants

The participants were 39 volunteers recruited on the Trinity College, University of Dublin campus, 22 men and 15 women (and 2 did not record their gender), aged 18 to 58 years, with an average age of 25 years.

Materials and design

Participants acted as their own controls and received four stories, one accidental purity violation, one accidental harm violation, one intentional purity violation, and one intentional harm violation. The contents of the two purity scenarios were about an agent eating a dog or an agent drinking urine, and the contents of the two harm scenarios were about an agent poisoning an individual or an agent causing an allergic reaction in an individual (adapted from Young & Saxe, 2011, with minor changes, see the Appendix). To control for content effects, we assigned the contents to the accidental and intentional conditions in two ways to create two sets, one set with the dog and allergy content as intentional and the urine and poison content as accidental, and another set with the dog and allergy content as accidental and the urine and poison content as intentional; participants were assigned at random to receive one or other set so that each participant received either the accidental or the intentional version of a scenario, and no participant received the same content for both. Each participant received the four scenarios in a different randomized order.

Our main measures were a moral responsibility and a moral wrongness judgment, which each participant completed: (a) "How morally responsible is Sam for what he did?", and they provided their judgments on a 7-point Likert-type scale with 1 anchored as not at all morally responsible, and 7 anchored as very morally responsible; and (b) "How morally wrong was what Sam did?" (1 = "not at all morally wrong", 7 = "very morally wrong"). They completed the measures in the fixed order of wrongness first and then responsibility. For the record, we also asked participants about some of the protagonist's emotions shame and guilt-as well as about moral character, and they completed the Moral Foundations Questionnaire (Graham et al., 2009). The results of the shame, guilt, and moral character measures did not show any systematic differences between the conditions, and for brevity we do not report them here.

Procedure

Participants were tested individually or in small groups. They were given a booklet with the four scenarios and the accompanying questions. They were instructed that the experiment was not a test of intelligence, and they were asked to read the scenarios carefully, to answer the questions in the order that they were given, and not to change any answers once they had completed them. They were asked not to turn to the next scenario until they had answered all questions for the current one. They completed the experiment in about 15 minutes.

Results and discussion

The data are archived (at https://reasoningand imagination.wordpress.com/). A repeated measures

analysis of variance (ANOVA) with the design of 2 (intent: intentional vs. accidental) $\times 2$ (domain: harm vs. purity) $\times 2$ (judgment type: wrongness vs. responsibility) on participants' ratings showed a main effect of intent, F(1, 38) = 142.98, MSE = 6.51, p < .001, $\eta_p^2 = .79$, as intentional violations were judged to be worse than accidental ones, a main effect of domain, F(1, 38) = 16.51, MSE = 2.38, p < .001, $\eta_p^2 = .30$, as harm violations were judged to be worse than purity ones, and no main effect of judgment type, F(1, 38) = 3.11, MSE = 2.08, p = .086. Intent and domain interacted, F(1, 38) =7.95, MSE = 2.08, p < .01, $\eta_p^2 = .17$, as Figure 1 shows. None of the other factors interacted, F < 1, and the three factors did not interact, F(1, 38) =2.66, MSE = 1.65, p = .11.

The four contrasts to decompose the interaction of intent and domain (with a Bonferroni corrected alpha of p < .01), showed that intentional violations were judged to be worse than accidental ones for harm, F (1, 38) = 168.71, MSE = 3.82, p < .001, $\eta_p^2 = .82$, and also for purity, F(1, 38) = 49.9, MSE = 6.33, p < .001, $\eta_p^2 = .57$ overall. Importantly, there was no difference between harm and purity for accidental violations, F < 1 (both about 2 on the 7-point scale), a very different finding from earlier results (Young & Saxe, 2011), as Figure 1 shows. Interestingly, intentional violations of harm were considered worse than intentional violations of purity, F(1, 38) = 17.55, MSE = 4.58, p < .001, $\eta_p^2 = .32$.

The experiment shows that intentional violations were judged worse than accidental ones regardless of whether they were harm or purity violations. The earlier suggestion that intentionality matters more for moral judgments of violations of harm than for those of violations of purity was based on an interaction in which intentional violations were judged equally harshly for harm and purity (6-7 on a 7-point scale, Young & Saxe, 2011), but accidental violations were judged more harshly for purity than for harm (5-6 vs. 1-2 on a 7-point scale, Young & Saxe, 2011). Our results also show an interaction between intentionality and domain, but the nature of the interaction is very different: Accidental violations were judged equally mildly for harm and purity (2-3 on a 7-point scale, for wrongness and responsibility), but intentional violations were judged more harshly for harm than for purity (6-7 versus 4-5 for wrongness and for responsibility), as Figure 1 shows. Judgments of both intentional and accidental purity violations



Figure 1. Moral responsibility and wrongness judgments for accidental and intentional harm and purity violations in Experiment 1. Error bars are standard error of the mean.

in our experiment were lower than earlier results (Young & Saxe, 2011).

We used the same vignettes as those in previous research, with minor adaptations, so the difference in the nature of the interaction is unlikely to result from differences in the materials. We can also rule out the possibility that the difference in the nature of the interactions arises from differences in culture, because the interaction observed here has been replicated with an American sample (Parkinson, 2015, Experiment 3). One conjecture is that the judgments in our experiment drew participants' attention to focus on the contribution of the agent-for example, "How morally responsible is Sam for what he did?" and "how morally wrong was what Sam did?"-compared to earlier judgments-for example, "how morally wrong was the action?" (Young & Saxe, 2011). We speculate that when participants' attention is drawn to the agent's contribution, they can calibrate their judgments to take into account whether the purity violation was intentional or accidental, just as they do for harm violations. The next experiment examines whether differences in the way people think about intentional and accidental harm and purity violations arise because of differences in their content and context, including in their target.

Experiment 2

The objective of the experiment was to test the idea that differences in judgments about moral domains, at least for harm and purity domains, result from differences in target, and hence are eliminated when content and context are controlled. The experiment aims first to control for potential content and context differences in harm and purity violations by constructing harm or purity versions of the same stories and, second, to examine the influence of the target, self or other, on judgments about an agent's moral responsibility and wrongness for intentional and accidental violations of harm and purity.

First, differences in judgments about harm and purity violations have been demonstrated in previous research for scenarios that contain very different contents. In this experiment we controlled for content and context by constructing scenarios that contained the same semantic and pragmatic material and were manipulated to be either harm or purity versions of the same story. For example, in one story, participants were told:

Joe and a co-worker are taking a tour of a chemical plant. During the coffee break, Joe goes to pour some coffee in the canteen.

In the harm version of the story they were told:

There is a tin of unmarked white powder on the little shelf beside the coffee urn. Joe believes that this is sugar, and does not know that this is dishwashing powder.

In the purity version of the story participants were told:

There is a plastic jar of milk in the little fridge beside the coffee urn. Joe believes that this is ordinary milk and does not know that this is one of the worker's breast milk for her baby.

In another story, participants were told:

Jane is a student on work placement with another student at a veterinary practice. During the afternoon break, Jane goes to the fridge to get something to drink. She sees a sealed plastic container with yellow liquid in it.

In the purity version of the story participants were told:

Jane believes that this is apple juice, and does not know that this is horse urine.

In the harm version of the story they were told:

Jane believes that this is apple juice, and does not know that this is horse medicine.

(and participants were informed that the medicine is poisonous). We adapted the four stories from the previous experiment to be either purity or harm versions to ensure that the only difference between the stories was whether they contained a purity or harm violation. Some moral violations may implicate both harm and purity principles, and our categorization of the stories as harm or purity violations refers to their primary moral domain. Each story was categorized based on our intuitions, as follows: It was categorized as a harm violation if it was about ingesting poisonous substances (poisonous horse medicine, poisonous star fruit); it was categorized as a purity violation if it was about ingesting disgusting substances (e.g., horse urine, dog meat). Experiment 1 has shown that intentional violations are judged to be worse than accidental ones for both harm and purity violations, and we predict that when content and context are controlled, any residual differences for harm and purity violations

will be eliminated, and intentional violations will be judged to be worse than accidental ones for both harm and purity violations.

Second, the target of harm violations is usually another person whereas the target of purity violations is usually the agent (e.g., Chakroff et al., 2013). We test the proposal that a violation that affects another person will be judged to be worse than one that affects only the agent, and that this difference will occur not only for harm violations but also for purity violations. The key manipulation was whether the target of the harm or purity violation was the agent or another person. For example, participants were told in the "self" version:

Now suppose that Joe knowingly puts this substance into his own coffee, but not his co-worker's.

Whereas they were told in the "other" version:

Now suppose that Joe knowingly puts this substance into his co-worker's coffee, but not his own.

(see the Appendix). We expect that violations that target others will be judged to be worse than ones that target the self. We also note that in the usual scenarios examined in previous research, agents often act as sole agents without accomplices in harm violations (e.g., Joe poisons his colleague), whereas they often act as joint agents with accomplices in purity violations (e.g., Sam takes a bag from the fridge containing his dog that *his partner* had put there). We took care to ensure that all of the scenarios in this experiment were constructed to contain only sole agents without accomplices.

Method

Participants

The participants were 22 undergraduates in Psychology at Trinity College Dublin, University of Dublin, who participated for research credits, 18 women and 4 men, aged 18 to 34 years. One other participant was eliminated for failure to complete most of the tasks.

Materials, design, and procedure

Participants acted as their own controls in a fully within-participant design. Each participant received four scenarios: one intentional harm violation, one accidental harm violation, one intentional purity violation, and one accidental purity violation. The scenarios contained four contents—about an agent putting a

substance in coffee at work, taking a substance from a vet's fridge, choosing an ingredient to cook for dinner, and choosing an ingredient for a meal on an exotic holiday. The four contents were assigned in a counterbalanced way to the four conditions for each participant, to ensure that the participant received a different content for each condition. Each participant was asked to imagine that the target of the violation was a person other than the agent-for example, "Now suppose that Joe knowingly puts this substance into his co-worker's coffee, but not his own"-and they completed a set of judgments about it; and then they were asked to imagine that the target of the violation was the agent-for example, "Now suppose that Joe knowingly puts this substance into his own coffee, but not his co-worker's"-and they completed a set of judgments about it. The order of supposing self or other as the target was randomized. The two primary judgments were (a) "How morally responsible is Sam for what he did?" (1 = not at all morally responsible, 5 = very morally responsible); (b) "How morally wrong is what Sam did?" (1 = "not at all morally wrong", 5 = "very morally wrong"). We also included measures of affect (disgust and anger) as a manipulation check to confirm that the purity violations elicited disgust, and the harm violations elicited anger, as shown in previous research (e.g., Horberg, Oveis, Keltner, & Cohen, 2009; Rozin, Lowery, Imada, & Haidt, 1999; also see Russell & Giner-Sorolla, 2011), and we report their results below. Additionally, participants completed judgments of cause and blame and the Moral Foundations Questionnaire, which for brevity we do not report here other than to note that the results for cause and blame measures did not show any systematic differences between the conditions. The procedure was the same as that in the previous experiment, and the materials were presented on the online platform SurveyGizmo (see http://www.surveygizmo.com).

Results and discussion

A repeated measures ANOVA with a design of 2 (target: self vs. other) × 2 (intent: accidental vs. intentional) × 2 (domain: harm vs. purity) × 2 (judgment type: wrongness vs. responsibility) on participants' ratings showed, first, that there was no main effect of domain, F(1, 21) = 1.1, MSE = 1.74, p = .31, and domain did not interact with intent, F(1, 21) = 2.29, MSE = 0.4, p = .15, nor with any other variable, F < 1. The absence of any interaction of domain (harm or

purity) with intent (accidental or intentional violation) is in stark contrast to the results of Experiment 1, as well as previous findings (Young & Saxe, 2011). We note that the experiment has sufficient power to detect such an interaction: A power analysis on the basis of the effect size for such an interaction in a study that controlled for target (Chakroff et al., 2013; $\eta_p^2 = .04$) indicates that an *n* of 22 participants has a statistical power of .79. The estimate is a conservative one; if it is based instead on the earlier study that found a domain by intent interaction (Young & Saxe, 2011; $\eta_p^2 = .23$ for harm and ingestion), the power to detect an interaction in this experiment increases to .99. It is also noteworthy that the results for the measures of affect (disgust and anger) confirmed that participants were more angered by the harm violations than by the purity ones (M = 2.6 vs. M = 2.43), and the difference was significant for violations of others, F(1, 21) = 8.25, MSE = 2.53, p = .009, $\eta_p^2 = .28$; they were more disgusted by the purity violations than by the harm ones (M = 2.84 vs. M = 2.68), and the effect was marginal for violations of the self, F(1,21) = 3.42, MSE = 1.46, p = .078, $\eta_p^2 = .14$. The result provides a manipulation check that participants indeed interpreted the harm and purity versions of the scenarios differently. The elimination of any effect of domain and any domain by intent interaction supports the suggestion that when the content and context of the violation are matched, including the target, participants do not judge intentional violations of harm to be worse than those of purity (pace Chakroff et al., 2013).

Second, the ANOVA showed a main effect of target, F(1, 21) = 39.43, MSE = 1, p < .001, $\eta_p^2 = .65$, as violations of others were judged worse than violations of the self; a main effect of intent, F(1, 21) = 214.6, MSE = 1.79, p < .001, $\eta_p^2 = .91$, as intentional violations were judged worse than accidental ones; and a main effect of judgment, F(1, 21) = 46.51, p < .001, MSE =0.91, $\eta_p^2 = .69$, as judgments of responsibility were higher than judgments of wrongness. Target interacted with intent, *F*(1, 21) = 21.11, *MSE* = 0.7, *p* < .001, $\eta_p^2 = .5$, and judgment type, F(1, 21) = 20.25, MSE =0.41, p < .001, $\eta_p^2 = .49$; target, intent, and judgment type interacted, F(1, 21) = 12, p < .01, MSE = 0.46, η_p^2 = .36, and there were no other interactions, F < 1, as Figure 2 shows. We carried out 12 contrasts to decompose the three-way interaction, with a Bonferroni corrected alpha of p < .004. It arises because intentional violations of the self were not judged to be as morally wrong as intentional violations of others



Figure 2. Moral responsibility and wrongness judgments for self and other targets for accidental and intentional violations of harm and purity in Experiment 2. Error bars are standard error of the mean.

(about 3 on the 5-point scale for self compared to almost 5 for others) F(1, 21) = 70.32, MSE = 0.84, p < .001, $\eta_p^2 = .77$, as Figure 2 shows, even though agents were judged to be morally responsible for intentional violations of the self as much as intentional violations of others (4–5 on the 5-point scale in each case) F < 1. There were no differences for accidental violations of the self or others, for wrongness judgments (1–2 on the 5-point scale), F(1, 21) = 6.43, MSE= 0.35, p = .019, $\eta_p^2 = .23$; or responsibility judgments (2–3 on the 5-point scale), F < 1. The result shows that the target of intentional moral violations, self or other, has a large influence on judgments about their wrongness.

Importantly, intentional violations were judged to be worse than accidental ones in every case—for judgments of the wrongness of violations of the self, F(1, 21) = 48.21, MSE = 0.94, p < .001, $\eta_p^2 = .7$, and others, F(1, 21) = 234.88, MSE = 0.71, p < .001, η_p^2 = .92, and for judgments of responsibility for violations of the self, F(1, 21) = 51.24, MSE = 1.6, p < .001, $\eta_p^2 = 71$, and others, F(1, 21) = 152.12, MSE = 0.73, p < .001, η_p^2 = .88. The result shows that intentional moral violations are judged more harshly than accidental ones, whether the target is the self or someone else. It is also interesting to note that violations were judged more harshly when participants judged the responsibility of the agent than when they judged the wrongness of the action: for accidental violations of others, *F* (1, 21) = 12.82, *MSE* = 0.7, *p* = .002, η_p^2 = .38, accidental violations of the self, *F*(1, 21) = 16.37, *MSE* = 0.76, *p* < .001, η_p^2 = .44, intentional violations of the self, *F* (1, 21) = 35.43, *MSE* = 0.97, *p* < .001, η_p^2 = .63, although not for intentional violations of others, *F*(1, 21) = 1.06, *MSE* = 0.39, *p* = .32, perhaps because of a ceiling effect in judgments of how wrong the latter are.

The experiment shows that intentional violations were judged worse than accidental ones regardless of whether they were harm or purity violations, consistent with the first experiment. Violations of others were judged worse than violations of the self. Although the target of a violation may contribute to whether it is considered as a harm or purity violation (Chakroff et al., 2013), it is important to note that the elimination of domain effects in scenarios that control for content and context indicates that there may be no overall effects of moral domain—at least for harm and purity domains—on moral judgments, but rather an effect of target.

General discussion

People judge that an agent who intentionally violates a moral principle has done something more morally wrong and is more morally responsible for it, than an agent who accidentally does so, regardless of whether it is a harm violation—for example, Jane gives poisonous horse medicine to another veterinary student—or a purity violation—for example, Jane gives horse urine to another veterinary student—as Experiments 1 and 2 both show. Importantly, participants judge accidental violations of purity to be as innocuous as accidental violations of harm (pace Young & Saxe, 2011), when they make judgments about the moral responsibility of the agent, or the wrongness of the agent's actions.

Violations that target another person—for example, Jane gives poisonous horse medicine to another veterinary student, or she gives horse urine to another veterinary student-are judged to be worse than violations that target the self-for example, Jane drinks poisonous horse medicine herself, or she drinks horse urine herself-as the second experiment shows. Harm violations tend to target others, and purity violations tend to target the self (see also Chakroff et al., 2013; Chakroff & Young, 2015b; Young & Tsoi, 2013), and differences in judgments about harm and purity are eliminated when their content and context are controlled, including their target. The intent behind a violation, and the target of the violation, appear to be the two factors that matter most in moral judgments about an agent's contribution, rather than the content of the violation as a harm violation or a purity one.

Reasoning about moral matters may be based on the same cognitive processes as reasoning about other contents (e.g., Bucciarelli et al., 2008; Cushman & Young, 2011; Stich, 2006; Sunstein, 2005; Uttich & Lombrozo, 2010). And reasoning about moral content may be prone to similar effects as reasoning about other contents (e.g., Rai & Holyoak, 2010; Shenhav & Greene, 2010). Reasoning about moral matters, like reasoning in general, may depend on the ability to envisage a situation, such as that Joe pours some powder into the coffee, and the events that led to it, such as Joe's beliefs and reasons for his actions, including his belief that the powder is sugar or that it is poison (e.g., Juhos et al., 2015; Walsh & Byrne, 2007). It may depend on an ability to consider alternative possibilities, such as that Joe did not pour the powder into the coffee, for example, when the outcome harms Joe or his colleague, and there may be a tendency to do so more often when Joe's action harms his colleague than when it harms Joe himself (e.g., Byrne, 2016). There may be many relevant factors in judgments of moral responsibility, such as the agent's identification with the behaviour (e.g., Cameron, Payne, & Knobe, 2010; Fischer & Ravizza, 2000; Knobe & Doris, 2010; Sripada, 2010; Woolfolk, Doris, & Darley, 2006) and inferences about their moral character (Pizarro & Tannenbaum, 2011). And the more severe the outcome, the more readily people assign responsibility to a causal agent (e.g., Walster, 1966; see also Trémolière & De Neys, 2013). Blame discounting following causal deviance or for agents whose attempts to harm fail but a similar harm occurs independently may reflect difficulty establishing the agent's contribution to the outcome (e.g., Cushman, 2008; Pizarro, Uhlmann, & Bloom, 2003). The experiments reported in this paper indicate that the cognitive processes that underpin the way people reason about the moral responsibility of an agent and the moral wrongness of an agent's actions for intentional and accidental violations are influenced by the way people think about violations of others compared to violations of the self. There may be no overall effects of moral domain, such as harm or purity, on moral judgments, but rather an effect of target.

Acknowledgements

We thank Maedhbh Etain Kelly, Jessica Dully, Jennifer Barry, and Michael Desmond Collins for help with coding the data. Some of the results were presented at the European Society for Cognitive Psychology conference in Budapest, 2013.

Funding

The research was supported by an Irish Research Council Government of Ireland Scholarship to Mary Parkinson; and by the Irish Research Council for the Humanities and Social Sciences.

References

- Anscombe, G. E. M. (1963). *Intention* (2nd ed.). Oxford: Basil Blackwell.
- Bratman, M. E. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, 3(2), 121–139.
- Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge: MIT Press.
- Byrne, R. M. J. (2016). Counterfactual thought. Annual Review of Psychology, 67, 135–157.
- Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, 23(4), 272–289.
- Carroll, R. (2015). Obama runs wild with Bear Grylls to promote action on climate change. Retrieved September 20, 2016, from https://www.theguardian.com/us-news/2015/dec/18/ barack-obama-bear-grylls-promote-action-climate-change

10 👄 M. PARKINSON AND R. M. J. BYRNE

- Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2015). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. Social Cognitive and Affective Neuroscience, 11, 476–84.
- Chakroff, A., Dungan, J., & Young, L. (2013). Harming ourselves and defiling others: What determines a moral domain? *PloS* one, 8(9), e74434.
- Chakroff, A., & Young, L. (2015a). Harmful situations, impure people: An attribution asymmetry across moral domains. *Cognition*, 136(issue), 30–37.
- Chakroff, A., & Young, L. (2015b). How the mind matters for morality. AJOB Neuroscience, 6(3), 43–48.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353–380.
- Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from non-moral psychological representations. *Cognitive Science*, 1, 24, 1052–1075.
- Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, *60*, 685–700.
- Fincham, F. D., & Jaspers, J. M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 13, pp. 82– 139). San Diego, CA: Academic Press.
- Fischer, J. M., & Ravizza, M. (2000). Responsibility and control: A theory of moral responsibility. Cambridge: Cambridge University Press.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046.
- Gray, K., & Schein, C. (2012). Two Minds Vs. Two philosophies: Mind perception defines morality and dissolves the debate between deontology and utilitarianism. *Review of Philosophy* and Psychology, 3(3), 405–423.
- Gubbins, E., & Byrne, R. M. (2014). Dual processes of emotion and reason in judgments about moral dilemmas. *Thinking & Reasoning*, 20(2), 245–268.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology*, 97(6), 963–976.
- Juhos, C., Quelhas, A. C., & Byrne, R. M. J. (2015). Reasoning about intentions: Counterexamples to reasons for actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41 (1), 55–76.
- Knobe, J. (2010). The person as moralist account and its alternatives. Behavioral and Brain Sciences, 33, 353–365.
- Knobe, J., & Doris, J. (2010). Responsibility. In J. Doris (Ed.), *The moral psychology handbook* (pp. 321–354). Oxford: Oxford University Press.
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. I. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, 37(6), 1036–1073.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, 25(2), 147–186.
- Miller, R. M., & Cushman, F. A. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral

condemnation of harm. *Social and Personality Psychology Compass*, 7(10), 707–718.

- Miller, R. M., Hannikainen, I., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, 14(3), 573–587.
- Nadelhoffer, T. (2004). On praise, side effects, and folk ascriptions of intentionality. *Journal of Theoretical and Philosophical Psychology*, 24(2), 196–213.
- Parkinson, M. (2015). Cognitive processes in moral judgment (Unpublished PhD thesis). Trinity College Dublin, University of Dublin, Ireland.
- Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgements of moral blame. In M. Mikulincer & P. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington: APA Press.
- Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39(6), 653–660.
- Rai, T. S., & Holyoak, K. J. (2010). Moral Principles or Consumer Preferences? Alternative Framings of the Trolley Problem. *Cognitive Science*, 34, 311–321.
- Rasga, C., Quelhas, A. C., & Byrne, R. M. (2016). Children's reasoning about other's intentions: False-belief and counterfactual conditional inferences. *Cognitive Development*, 40, 46–59.
- Rottman, J., Kelemen, D., & Young, L. (2014). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition*, 130(2), 217–226.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, *76*(4), 574–586.
- Russell, P. S., & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion*, 11(2), 233–240.
- Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review*, 101(4), 632–652.
- Shaver, K. G. (1985). The attribution of blame: Causality, responsibility, and blameworthiness. New York: Springer-Verlag.
- Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, 67, 667–677.
- Shoemaker, D. (2007). Moral address, moral responsibility, and the boundaries of the moral community. *Ethics*, *118*(1), 70–108.
- Shultz, T. R., & Schleifer, M. (1983). Towards a refinement of attribution concepts. In J. Jaspers, F. D. Fincham, & M. Hewstone (Eds.), Attribution theory and research: Conceptual, developmental and social dimensions (pp. 37–62). New York: Academic Press.
- Shultz, T. R., & Wright, K. (1985). Concepts of negligence and intention in the assignment of moral responsibility. *Canadian Journal of Behavioural Science/Revue canadienne* des sciences du comportement, 17(2), 97–108.
- Shultz, T. R., Wright, K., & Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Development*, 57, 177–184.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations, of suffering. In A. Brandt &

P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York: Routledge.

- Sripada, C. S. (2010). The deep self model and asymmetries in folk judgments about intentional action. *Philosophical Studies*, 151 (2), 159–176.
- Stich, S. (2006). Is morality an elegant machine or a kludge?. *Journal of Cognition and Culture, 6*(1–2), 181–189.
- Sunstein, C. R. (2005). Moral Heuristics. Behavioral and Brain Sciences, 28, 531–573.
- Sverdlik, S. (2004). Intentionality and moral judgments in commonsense thought about action. *Journal of Theoretical and Philosophical Psychology*, 24(2), 224–236.
- Trémolière, B., & De Neys, W. (2013). Methodological concerns in moral judgment research: Severity of harm shapes moral decisions. *Journal of Cognitive Psychology*, 25(8), 989–993.
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116(1), 87–100.
- Von Wright, G. H. (1983). Practical reason. Oxford: Basil Blackwell.

Walsh, C. R., & Byrne, R. M. (2007). How people think "if only ... " about reasons for actions. *Thinking & Reasoning*, 13(4), 461–483.

- Walster, E. (1966). Assignment of responsibility for an accident. Journal of Personality and Social Psychology, 3, 73–79.
- Watson, B. C. S. (1999). The Western ethical tradition and the morality of the warrior. Armed Forces and Society, 26, 55–72.

Weiner, B. (1995). Judgments of responsibility: A foundation for a theory of social conduct. New York: Guilford Press.

- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100(2), 283–301.
- Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, 120(2), 202–214.
- Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. Social and Personality Psychology Compass, 7(8), 585–604.

Appendix

Stories used in the experiments

Experiment 1

The wording for the intentional version is given, with the wording for the accidental version in italics.

1. Purity: dog. Imagine that a car just killed Sam's dog, mangling her badly. Sam's partner collects the remains and freezes them. Later that week, Sam decides to make himself dinner. He sees several packages in the freezer. They are correctly/ incorrectly labelled "dog" and it is the meat from his dog. Knowingly/without realizing it, Sam ends up eating his dog for dinner.

2. Purity: urine. Imagine that Jane is waiting to brush her teeth while her friend is in the bathroom. As her friend leaves, Jane notices a cup and a pregnancy test. Jane's friend tells her/forgets to tell her that the cup contains her friend's urine for her pregnancy test. Jane finishes brushing her teeth and uses that cup to rinse out her mouth. Knowingly/without realizing it, Jane rinses out her mouth with her friend's urine.

3. Harm: allergy. Imagine Ann's cousin is over for dinner. Ann knows/does not know her cousin is allergic to peanuts. Ann cooks dinner following a recipe from her favourite cookbook closely, which contains peanuts. She grinds up the peanuts, adds them in, and serves her cousin.

4. Harm: poison. Imagine Joe and a co-worker are taking a tour of a chemical plant. During the coffee break, Joe goes to pour some coffee. Joe likes his black, but his co-worker asks for sugar in his. He knows/does not know that he is putting poison in her coffee. Joe spoons some powder into his co-worker's coffee.

Experiment 2

The wording for the intentional version is given, with the wording for the accidental version in italics. The purity version is given first then the harm version, and the self version is given and then the other version.

1. Coffee. Joe and a co-worker are taking a tour of a chemical plant. During the coffee break, Joe goes to pour some coffee in the canteen.

Purity. There is a plastic jar of milk in the little fridge beside the coffee urn. Joe knows that this is one of the worker's breast milk for her baby, and that this is not ordinary milk/*Joe believes that this is ordinary milk, and does not know that this is one of the worker's breast milk for her baby.*

Harm. There is a tin of unmarked white powder on the little shelf beside the coffee urn. Joe knows that this is dishwashing powder, and that this is not sugar/*Joe believes that this is sugar, and does not know that this is dishwashing powder.*

Self. Now suppose that knowingly/*accidentally* Joe puts this substance* into his own coffee, but not his co-worker's.

Other. Now suppose that knowingly/*accidentally* Joe puts this substance* into his co-worker's coffee, but not his own.

(*"poisonous substance" was used in the harm version).

2. Juice. Jane is a student on work placement with another student at a veterinary practice. During the afternoon break, Jane goes to the fridge to get something to drink. She sees a sealed plastic container with yellow liquid in it.

Purity. Jane believes that this is apple juice, and does not know that this is horse urine/*Jane knows that this is horse urine, and that this is not apple juice.*

Harm. Jane believes that this is apple juice, and does not know that this is horse medicine/*Jane knows that this is horse medicine, and that this is not apple juice.*

Self. Now suppose that knowingly/*accidentally* Jane pours a glass of the liquid* for herself, but not for the other student.

Other. Now suppose that knowingly/*accidentally* Jane pours a glass of the liquid* for the other student, but not for herself.

(*"poisonous liquid" was used in the harm version).

3. Dinner. Ann's friend Sinead is visiting her for dinner one night. That morning, Ann goes out to the woods and handpicks some mushrooms for dinner. She decides to cook roast beef with stuffed roasted mushrooms as a side dish.

Purity. Ann knows that these are hallucinogenic "magic" mushrooms, and that these are not wild "morel" mushrooms/*Ann* believes that these are wild "morel" mushrooms, and does not know that these are hallucinogenic "magic" mushrooms.

Harm. Ann knows that these are poisonous "lorchel" mushrooms, and that these are not wild "morel" mushrooms/Ann believes that these are wild "morel" mushrooms, and does not know that these are poisonous "lorchel" mushrooms.

Self. Now suppose that knowingly/*accidentally* Ann serves these mushrooms* to herself, but not to Sinead.

Other. Now suppose that knowingly/*accidentally* Ann serves these mushrooms* to Sinead, but not to herself.

(*"hallucinogenic mushrooms" was used in the purity version and "poisonous mushrooms" in the harm version).

4. Holiday. Sam is on holiday with a friend in Southeast Asia. One evening, he decides to eat in a local restaurant he passes by.

Purity. He sees a meal on the menu, which contains a meat that looks like beef. Sam knows that this meat is dog, and that this is not beef/*Sam believes that this meat is beef, and does not know that this meat is dog.*

Harm. He sees a meal on the menu, which contains a fruit that looks like papaya. Sam knows that this is star fruit, which contains poisonous oxalic acid, and that this is not papaya/*Sam believes* that this fruit is papaya, and does not know that this fruit is star fruit, which contains poisonous oxalic acid.

Self. Now suppose that knowingly/*accidentally* Sam orders this meal* for himself, but not to bring back to the hotel for his friend.

Other. Now suppose that knowingly/*accidentally* Sam orders this meal* to bring back to the hotel for his friend, but not for himself.

(*"this meal containing dog" was used in the purity version and "this meal containing poisonous starfruit" in the harm version).