



# When ignorance is no excuse: Different roles for intent across moral domains

Liane Young<sup>a,b,\*</sup>, Rebecca Saxe<sup>a</sup>

<sup>a</sup> Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, MA, USA

<sup>b</sup> Department of Psychology, Boston College, MA, USA

## ARTICLE INFO

### Article history:

Received 5 October 2010

Revised 12 April 2011

Accepted 25 April 2011

Available online 23 May 2011

### Keywords:

Morality

Intent

Disgust

Purity

Incest

Accident

## ABSTRACT

A key factor in legal and moral judgments is intent. Intent differentiates, for instance, murder from manslaughter. Is this true for all moral judgments? People deliver moral judgments of many kinds of actions, including harmful actions (e.g., assault) and purity violations (e.g., incest, consuming taboo substances). We show that intent is a key factor for moral judgments of harm, but less of a factor for purity violations. Based on the agent's innocent intent, participants judged accidental harms less morally wrong than accidental incest; based on the agent's guilty intent, participants judged failed attempts to harm more morally wrong than failed attempts to commit incest. These patterns were specific to moral judgments versus judgments of the agent's control, knowledge, or intent, the action's overall emotional salience, or participants' ratings of disgust. The current results therefore reveal distinct cognitive signatures of distinct moral domains, and may inform the distinct functional roles of moral norms.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

If a hunter sees a bird, aims and fires, but hits his hunting partner, he is generally forgiven for the accident, and legal penalties are minimal. By contrast, if the hunter aims and fires at his hunting partner, ordinary people and legal practice agree that the act was wrong and the actor deserves punishment. Thus, a key factor in legal and moral judgments – and the difference between murder and manslaughter – is the agent's intent (Mikhail, 2007). Indeed, among actions that cause harm, intentional actions are judged worse than accidents, even accidents with serious negative consequences (Cushman, 2008; Knobe, 2005; Malle & Knobe, 1997; Piaget, 1965/1932; Singer, Kiebel, Winston, Dolan, & Frith, 2004; Young, Cushman, Hauser, & Saxe, 2007). After young children develop the capacity to reason about the minds of others, they too start to differentiate between intentional harms and accidents (Baird &

Astington, 2004; Killen, Lynn Mulvey, Richardson, Jampol, & Woodward, 2011). The law reflects this intuition as well: *mens rea* or “guilty mind” is often a key element of a crime (Hart, 1968). Yet, as we explore in the present paper, the simple rule that “intent matters” may not be true for all moral judgments.

People make moral judgments of many different kinds of actions, including harming or taking unfair advantage of others, sleeping with blood relatives, and ingesting taboo substances. One proposal is that moral judgments of these different actions reflect distinct moral domains (Blair, Marsh, Finger, Blair, & Luo, 2006; Graham & Haidt, 2010; Haidt, 2007; Rai & Fiske, 2011; Rozin, Lowery, Imada, & Haidt, 1999; Sinnott-Armstrong, 2008; Walker & Hennig, 2004; Wright & Baril, 2011). For example, norms against harming others may belong to one domain (“Harm”), while norms against incest or ingesting taboo substances (i.e. pathogen ingestion) belong to another domain (“Purity”) (Haidt, 2007).

Recent research suggests that judgments across distinct domains (e.g., Harm, Purity) are associated with different patterns of emotional responding (Monin, Pizarro, & Beer, 2007; Rozin et al., 1999), neural activity (Borg, Hynes,

\* Corresponding author at: Department of Psychology, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Tel./fax: +1 617 552 4102.

E-mail address: [liane.young@bc.edu](mailto:liane.young@bc.edu) (L. Young).

Van Horn, Grafton, & Sinnott-Armstrong, 2006; Moll et al., 2005), and behavioral judgment (Haidt, 2007; Haidt, Koller, & Dias, 1993; Wright & Baril, 2011). For example, witnessing one person harming another tends to elicit anger (Blair et al., 2006; Kedia, Berthoz, Wessa, Hilton, & Martinot, 2008), whereas purity violations tend to elicit disgust (Moll et al., 2005; Rozin et al., 1999; Schaich Borg, Lieberman, & Kiehl, 2008). Different moral domains are emphasized to different degrees across communities; in general, political liberals assign greater moral weight to harm than to purity, while political conservatives assign similar moral weight to both domains (Graham & Haidt, 2010; Graham, Haidt, & Nosek, 2009; Haidt, 2007).

An interesting behavioral difference is that whereas people find it easy to defend their moral judgments of harmful actions, by appealing to the experience of the victim, people find it harder to justify their condemnation of victimless purity violations like consensual incest (Haidt, 2001; Haidt et al., 1993). People often initially claim that purity violations involve harm to a victim (Royzman, Lee-man, & Baron, 2009), but even when assured that no harm occurred many still insist that purity violations are ‘simply wrong’ even though they cannot explain why – “moral dumbfounding” (Haidt, 2001). As a result, justifying judgments of purity violations appears to be more effortful than justifying judgments of harms (Wright & Baril, 2011).

A further prediction of the proposal that harm and purity reflect distinct moral domains is that moral judgments of these two kinds of violations rely on distinct cognitive computations, or at least differently weighted cognitive processes (Monin et al., 2007). The current study focuses on one possible difference between moral domains: the moral weight assigned to the agent’s intent. We hypothesize that the role of intent in moral judgment depends on the kind of moral violation under evaluation. While the key role of intent has been established in moral judgments of harm, intent may play a significantly smaller role in moral judgments of purity violations, including incest and ingesting taboo substances.

Our hypothesis was initially motivated by an observation: in at least some cultures, people who commit purity violations accidentally and unknowingly are nevertheless considered impure and immoral (J. Henrich, personal communication). We asked whether a similar pattern of moral judgments would be observed among Western adults, who typically treat intent as the primary factor in moral judgments of harm violations (Cushman, 2008; Mikhail, 2007; Young et al., 2007). If so, the difference may lie in the fact that purity violations are often “victimless” transgressions: aside from the agents themselves, no one else is directly affected by the consumption of taboo substances or consensual incest (Rozin et al., 1999). Considering someone’s intent may be more important for regulating behavior that directly affects other people (Baumeister, Stillwell, & Heatherton, 1994; Kedia et al., 2008). Indeed, in paradigmatic cases of harm, at least one person harms at least one other person (Gray & Wegner, 2009; Royzman et al., 2009; Turiel, 1983; Waytz, Gray, Epley, & Wegner, 2010). The victim, or the judge, in a court of law, may then demand an explanation from the perpetrator, and the perpe-

trator might appeal to his or her innocent intentions (e.g., it was an accident).

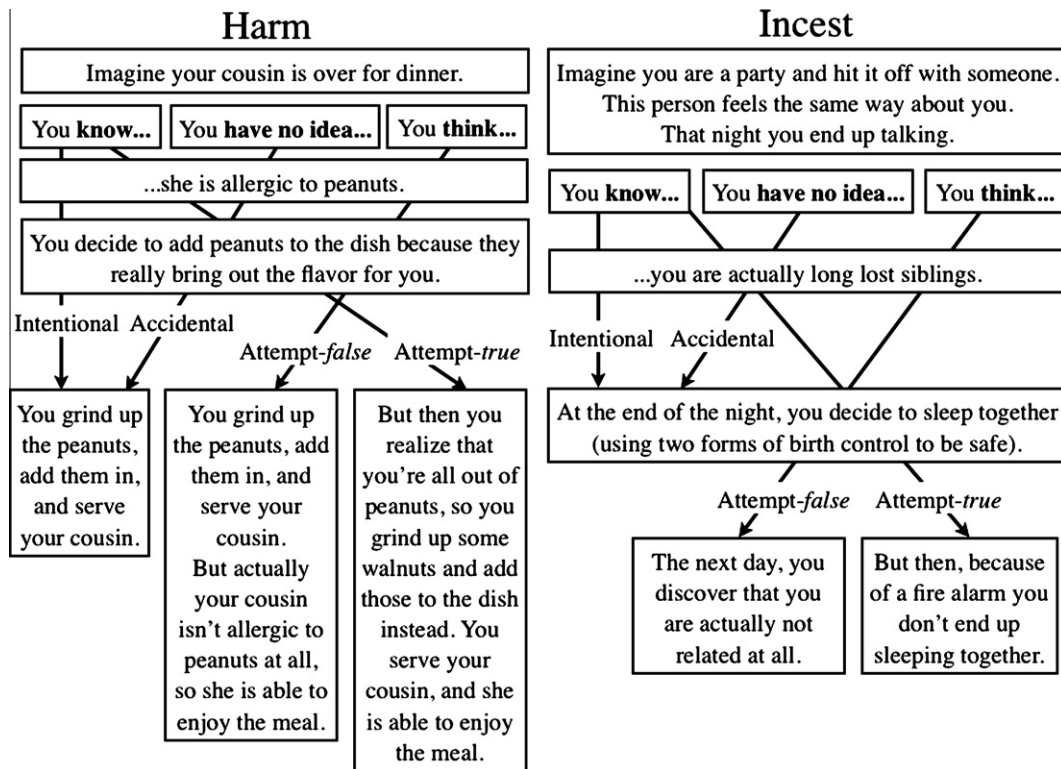
In the current study, we investigate whether (both innocent and guilty) intentions would matter more for moral judgments of harm and less for moral judgments of purity violations. Experiment 1 tests the hypothesis that accidental harm would be judged less morally wrong than accidental incest or accidental pathogen ingestion, based on false beliefs and innocent intentions. Experiment 2 determines that differences in participants’ moral judgments of harm versus purity violations cannot be attributed to differences in their ratings of overall emotional salience, control, knowledge, or intent. Experiment 3 determines that participants’ moral judgments do not simply reflect participants’ expressions of disgust in response to the stimuli. Experiment 4 tests the hypothesis that failed attempts to harm are judged more morally wrong than failed attempts to commit incest, based on false beliefs and guilty intentions. Experiment 5 investigates two distinct kinds of failed attempts to further explore potential differences in the cognitive processes for moral judgments of harm versus incest.

These experiments aim to show that moral judgments of harm violations depend more on intent, and moral judgments of purity violations depend more on the outcome, actual or anticipated. The current study thus characterizes the differences in the cognitive processes for moral judgments across two moral domains: harm versus purity.

## 2. Experiment 1: intentional versus accidental violations

We hypothesized that participants would assign more moral weight to the agent’s innocent intention when judging accidental harms, compared to accidental purity violations. Specifically, we hypothesized that, for violations that are judged equally morally wrong when committed intentionally (i.e., harm, incest, ingesting taboo substances), when those same actions occur by accident, accidental *harms* would be judged more leniently than accidental *purity* violations. For example, intentional poisoning and intentional incest might be judged equally morally wrong, but accidental poisoning would be judged less morally wrong than accidental incest.

We tested this hypothesis in two experiments, using moral scenarios described in the second person (“You meet someone at a party...”, Experiment 1A), or in the third person (“Sam meets someone at a party...”, Experiment 1B). We reasoned that moral judgments pertaining to purity norms might be more strongly invoked by stimuli in the second person, which highlights the self-relevance of the action. It is possible that purity norms apply more often to one’s own actions (e.g., we care about not eating taboo foods ourselves, and we care about avoiding our own siblings as sexual partners), while harm norms are commonly applied in third person contexts as well (e.g., we evaluate both our own and others’ actions on the basis of whether anyone is hurt). Nevertheless, general structural features of moral judgments in specific domains should apply to both second and third person contexts and so should be observable for both kinds of stimuli. Therefore for both



**Fig. 1.** Schematic of sample scenarios for harm (left) and purity (right). In Experiment 1, participants judge intentional and accidental violations. In Experiments 4 and 5, participants additionally judge attempted violations: true belief/failed act, false belief/completed act.

second and third person stimuli, we hypothesized a reduced role for intent in moral judgments of purity versus harm violations.

## 2.1. Method

### 2.1.1. Experiment 1A

We collected data from 262 participants (134 female, aged 18–68 yrs, mean = 37 yrs, standard deviation = 12 yrs), using Amazon Mechanical Turk (<http://www.mturk.com/>). Participants were paid \$0.10 for approximately one minute of time. Three measures were taken to screen out repeat participants: (1) we asked that people not participate if they had previously taken a similar survey, (2) participants answered a final question about whether they had completed a similar survey before and, if so, its topic, (3) we eliminated data from participants with identical worker IDs. We eliminated 21 participants, yielding a total of 241 participants and an approximately equal number of participants per condition.

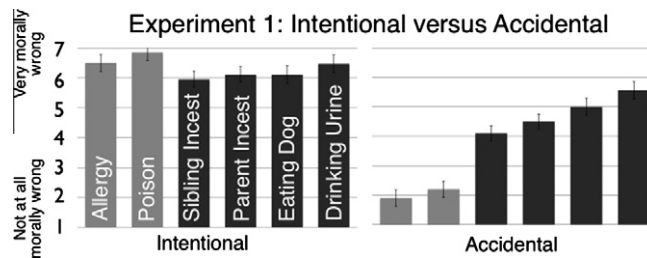
Each participant made a moral judgment for a single scenario. Participants were assigned randomly to one of six conditions in a 2 (intentional versus accidental)  $\times$  3 (harm versus incest versus ingestion) between-subjects experimental design (Fig. 1; for full text of all scenarios, see [Supplementary Material](#)). In Experiment 1A, moral scenarios were presented in the second person (e.g., “Your cousin comes over for dinner”, see Fig. 1). Participants

judged the moral wrongness of the action, on a 7-point scale anchored at “not at all morally wrong” (1) to “very morally wrong” (7). Two scenarios represented each domain: harm (allergy,<sup>1</sup> poison), incest (parent, sibling) or ingestion (dog meat, urine). There were two versions of each scenario (intentional, accidental); each participant saw only one version.

### 2.1.2. Experiment 1B

We collected data from 80 new participants, using Amazon Mechanical Turk, as in Experiment 1A, to investigate whether the key effects would replicate using *third-person* moral scenarios. Experiment 1B used a subset of the same moral scenarios (allergy, sibling incest) but presented them in the third person, replacing “you” with the name of a specific individual (e.g., “Sam”). As in Experiment 1A, each participant made a moral judgment for a single scenario, in a 2 (intentional versus accidental)  $\times$  2 (harm versus incest) between-subjects experimental design. Participants judged the moral wrongness of the action, on a 7-point scale anchored at “not at all morally wrong” (1) to “very morally wrong” (7).

<sup>1</sup> We tested two different versions of the allergy scenario: (1) one version that explicitly stipulated the allergy as fatal and (2) one version that did not. We found no differences between versions in either the intentional condition ( $t(38) = 0.85, p = 0.4$ ) or accidental ( $t(38) = 0.71, p = 0.5$ ) condition. The analyses use version 2.



**Fig. 2.** Experiment 1: Intentional versus accidental violations. Moral judgments for harm versus purity violations. Participants judged two harm scenarios (light bars) and four purity scenarios (dark bars), describing intentional (left) and accidental (right) violations. Moral judgments of intentional (dark) and accidental (light) violations of harm (left), incest (center), and ingestion (right) norms. Error bars represent standard error.

## 2.2. Results and discussion

### 2.2.1. Experiment 1A

To verify that judgments of the two stories representing each violation type did not differ, we conducted three 2 (story)  $\times$  2 (intent) ANOVAs: for harm, incest, and ingestion. These analyses yielded the same pattern: a main effect of intent (i.e., intentional violations were judged morally worse than accidental violations) but not story (i.e., stories within a domain were not judged differently), and no story by intent interaction, indicating no differences in the role of intent between stories within a domain.

<sup>2</sup> Therefore, the analyses below collapse across stories.

We conducted three 2 (intent)  $\times$  2 (domain) ANOVAs, allowing us to compare the role of intent for (1) harm versus incest, (2) harm versus ingestion, and (3) incest versus ingestion.<sup>3</sup> In all cases, intentional violations were judged morally worse than accidental violations, as indicated by a main effect of intent. (1) Comparing harm to incest, we observed main effects of intent ( $F(1, 169) = 148.1, p < 0.001$ , partial  $\eta^2 = 0.47$ ), domain ( $F(1, 169) = 8.5, p = 0.004$ , partial  $\eta^2 = 0.05$ ), and the critical intent by domain interaction ( $F(1, 169) = 29.1, p < 0.001$ , partial  $\eta^2 = 0.15$ ), indicating a different role for intent for judgments of harm versus incest. (2) Comparing harm to ingestion, we also observed main effects of intent ( $F(1, 150) = 106.7, p < 0.001$ , partial  $\eta^2 = 0.42$ ), domain ( $F(1, 150) = 27.1, p < 0.001$ , partial  $\eta^2 = 0.16$ ), and the same intent by domain interaction ( $F(1, 150) = 44.0, p < 0.001$ , partial  $\eta^2 = 0.23$ ), indicating a different role for intent for judgments of harm versus ingestion. (3) However,

<sup>2</sup> (1) For harm, we found a significant main effect of intent ( $F(1, 79) = 192.9, p < 0.001$ , partial  $\eta^2 = 0.72$ ), but no effect of story ( $F(1, 79) = 0.95, p = 0.33$ ) or story by intent interaction ( $F(1, 79) = 0.01, p = 0.94$ ). (2) For incest, we found a significant main effect of intent ( $F(1, 89) = 18.3, p < 0.001$ , partial  $\eta^2 = 0.18$ ), but no effect of story ( $F(1, 89) = 0.51, p = 0.48$ ) or story by intent interaction ( $F(1, 89) = 0.12, p = 0.73$ ). (3) For pathogen, we found a significant main effect of intent ( $F(1, 70) = 5.16, p = 0.03$ , partial  $\eta^2 = 0.07$ ), but no effect of story ( $F(1, 70) = 1.1, p = 0.31$ ) or story by intent interaction ( $F(1, 70) = 0.05, p = 0.83$ ). Independent-samples *t*-tests revealed a difference between intentional and accidental actions in both harm stories ( $p$ 's  $< 0.001$ ), both incest stories ( $p$ 's  $< 0.001$ ), and a marginal difference in both ingestion stories ( $p$ 's  $> 0.11$ ).

<sup>3</sup> An overall 2 (intentional versus accidental)  $\times$  3 (harm versus incest versus ingestion) ANOVA yielded a main effect of intent ( $F(1, 240) = 117.8, p < 0.001$ , partial  $\eta^2 = 0.33$ ), a main effect of domain ( $F(1, 240) = 12.4, p < 0.001$ , partial  $\eta^2 = 0.10$ ), and the critical intent by domain interaction ( $F(1, 240) = 22.9, p < 0.001$ , partial  $\eta^2 = 0.16$ ), indicating a different role for intent depending on the domain (Fig. 2).

comparing incest to ingestion, we observed main effects of intent ( $F(1, 160) = 21.8, p < 0.001$ , partial  $\eta^2 = 0.12$ ), domain ( $F(1, 160) = 4.7, p = 0.03$ , partial  $\eta^2 = 0.03$ ), but no intent by domain interaction ( $F(1, 160) = 1.7, p = 0.20$ , partial  $\eta^2 = 0.01$ ). Together, these results suggest significant differences in the role of intent for harm violations, compared to both kinds of purity violations (i.e., incest, ingestion).

Critically, independent-samples *t*-tests revealed the key predicted differences in the case of accidents. Incest was judged morally worse than harm when both were committed accidentally (accidental incest: 4.24 out of 7; accidental harm: 2.05;  $t(88) = 5.0, p < 0.001$ ), though harm was judged slightly worse than incest when both were committed intentionally (intentional incest: 6.03; intentional harm: 6.68;  $t(78) = 2.5, p = 0.02$ ). Similarly, ingesting taboo substances was judged morally worse than harm when both were committed accidentally (accidental ingestion: 5.28;  $t(74) = 6.8, p < 0.001$ ), but not significantly different when both were committed intentionally (intentional ingestion: 6.29;  $t(73) = 1.5, p = 0.15$ ).

### 3. Experiment 1B

Replicating the results of Experiment 1A, a 2 (domain: harm versus incest)  $\times$  2 (intent: bad versus neutral) between-subjects ANOVA revealed a main effect of intent (i.e., intentional violations were judged morally worse than accidental violations,  $F(1, 76) = 40.8, p < 0.001$ , partial  $\eta^2 = 0.35$ ), no main effect of domain ( $F(1, 76) = 2.8, p = 0.10$ , partial  $\eta^2 = 0.04$ ), and the key predicted interaction between domain and intent ( $F(1, 76) = 5.6, p = 0.02$ , partial  $\eta^2 = 0.07$ ). As in Experiment 1A, incest was judged morally worse than harm when both were committed accidentally (accidental incest: 3.86; accidental harm: 2.09;  $t(49) = 3.2, p = 0.002$ ), and there was no difference when both were committed intentionally (intentional incest: 5.64, intentional harm: 5.94;  $t(27) = 0.49, p = 0.63$ ).

Combining the results of Experiments 1A and 1B, a 2 (perspective: second versus third person)  $\times$  2 (domain)  $\times$  2 (intent) between-subjects ANOVA revealed no effects or interactions involving perspective ( $p$ 's  $> 0.4$ ). As in the separate analyses of Experiments 1A and 1B, this combined analysis did reveal a main effect of intent ( $F(1, 162) = 112.7, p < 0.001$ , partial  $\eta^2 = 0.41$ ) and the key interaction between domain and intent ( $F(1, 162) = 17.7, p < 0.001$ , partial  $\eta^2 = 0.10$ ). Because there was no effect of

perspective in Experiment 1, subsequent experiments were conducted using only second-person stimuli.

Together, the results of Experiments 1A and 1B support the hypothesis that innocent intent plays a greater role in moral judgments of harm versus purity violations. As predicted, accidental harms were judged less morally wrong than accidental purity violations, though harms were not judged less morally wrong than purity violations in general.

#### 4. Experiment 2: emotion, control, knowledge, and intent

In Experiment 1, participants judged accidental purity violations more morally wrong than accidental harms. Could this pattern of moral judgments be accounted for by other perceived differences between accidental purity violations and accidental harms? Participants might judge accidental incest morally worse if they find accidental incest to be more emotionally upsetting overall compared to accidental harm. Alternatively, participants might judge accidental incest morally worse if they perceive accidental incest (compared to accidental harm) to be committed with greater control of the situation, greater knowledge, or greater intent (e.g., participants might not accept that they could commit incest without knowing it).

Experiment 2 therefore aimed to determine whether the difference in participants' moral judgments of harm versus purity violations could be due to differences in participants' ratings of the overall emotional salience of the harm versus incest scenarios, the agent's control, the agent's knowledge, or the agent's intent. We hypothesized that participants would not perceive differences along any of these dimensions for harm versus incest scenarios. Experiment 2 also provided a replication of the moral judgment pattern obtained in Experiment 1.

##### 4.1. Method

We collected data from 320 new participants, using Amazon Mechanical Turk, as in Experiment 1. Separate groups of 80 participants made each of the following judgments for intentional and accidental incest (sibling) and harm (poison) scenarios: (1) How *emotionally upset* did you feel while reading this story? (1 = not at all, 7 = very much); (2) Did you have *control* over whether or not you [slept with your sibling/poisoned your co-worker]? (1 = definitely not, 7 = definitely); (3) Did you *know* you were [sleeping with your sibling/poisoning your co-worker]? (1 = definitely not, 7 = definitely); (4) Did you [sleep with your sibling/poison your co-worker] *intentionally*? (1 = definitely not, 7 = definitely); (5) How *morally wrong* was the action? (1 = not at all, 7 = very much).

##### 4.2. Results and discussion

We conducted separate 2 (domain: harm versus incest)  $\times$  2 (intent: bad versus neutral) between-subjects ANOVAs and independent-samples *t*-tests for each kind of non-moral judgment: emotion, control, knowledge,

and intent (Table 1). As predicted, we found no difference between harm and incest along any of these non-moral dimensions, for either accidental or intentional violations.

First, accidental incest and accidental harm were judged equally emotionally upsetting. For emotion judgments, we found no main effect of intent ( $F(1, 76) = 1.2, p = 0.275$ , partial  $\eta^2 = 0.02$ ), no main effect of domain ( $F(1, 76) = 2.99, p = 0.09$ , partial  $\eta^2 = 0.038$ ), and no interaction between domain and intent ( $F(1, 76) = 0.62, p = 0.43$ , partial  $\eta^2 = 0.008$ ). There was no difference in the emotional salience of harm versus incest for accidental (harm: 4.50 out of 7, incest: 4.05;  $t(38) = 0.66, p = 0.55$ ) or intentional actions (harm: 5.40; incest: 4.20;  $t(38) = 1.81, p = 0.08$ ).

Second, accidental incest and accidental harm were perceived as equivalent in terms of the agent's control over the outcome, knowledge of the situation, and intention to commit a violation. As a manipulation check, for both harm and incest, we also determined that intentional actions were judged as having been performed with more control, knowledge, and intent, compared to accidental actions. Therefore, for all three judgments (control, knowledge, intent), we found a main effect of intent (intentional > accidental, all  $p$ 's < 0.001), no main effect of domain (harm = incest, all  $p$ 's > 0.49), and no domain by intent interaction (all  $p$ 's > 0.37).

Finally, we replicated the results of Experiments 1A and 1B for moral judgment: the same domain by intent interaction pattern. We found the same main effect of intent (i.e., intentional violations were judged morally worse than accidental violations,  $F(1, 76) = 46.9, p < 0.001$ , partial  $\eta^2 = 0.381$ ), no main effect of domain (i.e., overall, incest was not judged morally worse than harm,  $F(1, 76) = 0.60, p = 0.44$ , partial  $\eta^2 = 0.008$ ), and the key predicted interaction between domain and intent ( $F(1, 76) = 15.0, p < 0.001$ , partial  $\eta^2 = 0.17$ ): as in the prior experiments, incest was judged morally worse than harm when both were committed accidentally (harm: 2.50; incest: 4.45;  $t(38) = 2.73, p = 0.01$ ), and harm was judged slightly morally worse than incest when both were committed intentionally (harm: 7.00; incest: 5.70;  $t(38) = 2.94, p = 0.01$ ).

In sum, participants did not judge incestuous acts to be more emotionally upsetting overall; in fact, intentional

**Table 1**

Experiments 2 and 3: Additional ratings (means, standard deviations). In Experiment 2 (rows 2–6), participants delivered ratings of the scenario's overall emotional salience, the agent's control, the agent's knowledge, the agent's intent, or the action's moral wrongness, for scenarios describing harm and incest, on a 7-point scale, anchored at "definitely not/not at all" (1) and "definitely/very much" (7). In Experiment 3 (row 1), participants rated the disgustingness of harm and purity violations on a 4-point scale, anchored at "not at all disgusting" (1) and "very disgusting" (4).

|               | Accidental  |             | Intentional |             |
|---------------|-------------|-------------|-------------|-------------|
|               | Harm        | Incest      | Harm        | Incest      |
| Disgust       | 1.34 (0.44) | 3.25 (0.44) | 2.19 (1.03) | 3.54 (0.36) |
| Emotion       | 4.50 (2.19) | 4.05 (2.14) | 5.40 (2.14) | 4.20 (2.07) |
| Control       | 3.15 (2.41) | 3.55 (2.50) | 5.85 (1.66) | 5.35 (2.21) |
| Knowledge     | 1.00 (0.00) | 1.35 (.10)  | 5.60 (2.46) | 5.90 (2.78) |
| Intent        | 2.20 (2.07) | 2.75 (2.20) | 5.15 (2.54) | 5.30 (2.16) |
| Morally wrong | 2.50 (2.28) | 4.45 (2.24) | 7.00 (0.00) | 5.70 (1.98) |

harms trended towards being judged as slightly more emotionally upsetting than intentional incest. Participants also judged incestuous and harmful acts, whether intentional or accidental, to be equally controllable, intentional, and the agents to possess equal knowledge of the situation.<sup>4</sup> Only moral judgments showed the key interaction pattern: incest was judged morally worse than harm, for accidents only.

### 5. Experiment 3: morality versus disgust

In Experiments 1 and 2, accidental incest was judged morally worse than accidental harm. However, in Experiment 2, accidental incest was judged no more emotionally upsetting, controllable, or intentional than accidental harm. Nevertheless, there remains another alternative hypothesis for why accidental incest is judged morally worse than accidental harm: perhaps participants are not making moral judgments of incestuous acts but instead using the moral judgment question as an opportunity to express their feelings of disgust at the scenarios. Moral violations frequently lead to feelings of disgust, and feelings of disgust can influence moral judgments. For example, after committing a moral violation, like cheating, people feel literally dirty – and choose a cleaning product as an experimental reward (Zhong & Liljenquist, 2006). Even young children use the term “disgusting” to describe moral violations (Danovitch & Bloom, 2009). Priming studies show that people who are feeling disgusted, because of a dirty environment (Schnall, Benton, & Harvey, 2008; Schnall, Haidt, Clore, & Jordan, 2008), a hypnotic suggestion (Wheatley & Haidt, 2005), or a bad taste in their mouth (Eskine, Kaciniak, & Prinz, 2011), make harsher moral judgments. Recent evidence also suggests that participants who are reminded of physical cleansing make harsher moral judgments of purity violations (and report being more politically conservative) (Helzer & Pizarro, 2011).

Since intentions might matter less for judging how disgusting versus how morally wrong an act is, we aimed to test whether our participants were actually expressing their feelings of disgust, rather than reporting their moral judgments. We therefore asked participants to provide both moral judgments and ‘disgust ratings’ of our stimuli. We reasoned that if the patterns of moral judgments differed from the patterns of disgust ratings across stimuli, this difference would provide evidence that participants (a) distinguish between questions about morality and questions about disgust, and (b) rely on different processes for establishing how disgusting an act is versus how morally wrong the same act is. Note, though, that these results would not contradict prior evidence that disgust plays a key role in moral judgments, and specifically in moral judgments about purity violations (Helzer & Pizarro, 2011; Inbar, Pizarro, & Bloom, 2009; Inbar, Pizarro, Knobe,

& Bloom, 2009). These results would serve only to show that disgust ratings and moral judgments, even for purity violations, are not one and the same.

#### 5.1. Method

We collected data from 160 new participants, using Amazon Mechanical Turk. Participants read a single moral scenario from one of four conditions (intentional harm, accidental harm, intentional incest, accidental incest). Afterwards participants made two judgments, using a 7-point scale: “how morally wrong” was the action, rated from “not at all morally wrong” (1) to “very morally wrong” (7), and “how disgusting” was the action, rated from “not at all disgusting” (1) to “very disgusting” (7). The order of the two judgments was counterbalanced across participants (moral judgment first versus disgust rating first).

#### 5.2. Results and discussion

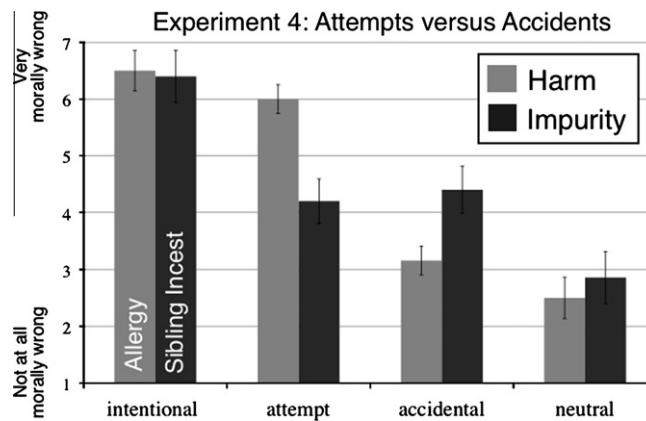
First, replicating the results of Experiments 1 and 2, a 2 (domain)  $\times$  2 (intent)  $\times$  2 (question order: moral judgment first versus disgust rating first) mixed-effects ANOVA of participants’ moral judgments yielded a main effect of intent (i.e., intentional violations were judged morally worse than accidental violations,  $F(1, 152) = 38.5$ ,  $p < 0.001$ , partial  $\eta^2 = 0.20$ ), and the predicted intent by domain interaction (i.e., intent matters more for harm than incest,  $F(1, 152) = 8.4$ ,  $p = 0.004$ , partial  $\eta^2 = 0.05$ ).<sup>5</sup> Accidental harm (mean: 2.93 out of 7) was judged less morally wrong than accidental incest (mean: 4.5;  $t(82) = -3.3$ ,  $p = 0.001$ ), and there was no difference between harm and incest when both were intentional (harm: 5.92; incest: 5.49;  $t(74) = 1.0$ ,  $p = 0.32$ ).

Second, we conducted a 2 (judgment type: how morally wrong versus how disgusting)  $\times$  2 (domain)  $\times$  2 (intent)  $\times$  2 (question order) mixed-effects ANOVA. This analysis revealed a main effect of judgment type ( $F(1, 152) = 5.7$ ,  $p = 0.02$ , partial  $\eta^2 = 0.04$ ); on average, the scenarios were judged more disgusting (5.0) than morally wrong (4.7). Critically, we observed the key predicted interactions involving judgment type (how morally wrong versus how disgusting): we found significant interactions between judgment and domain ( $F(1, 152) = 4.8$ ,  $p = 0.03$ , partial  $\eta^2 = 0.03$ ) and judgment and intent ( $F(1, 152) = 8.0$ ,  $p = 0.005$ , partial  $\eta^2 = 0.05$ ).<sup>6</sup> These key interactions indicate that participants’ answers to the questions “how morally wrong” versus “how disgusting” an act is are influenced differently by the factors of domain (i.e., harm versus incest) and intent (i.e., intentional versus accidental). To unpack these interactions, we conducted independent-samples *t*-tests. Participants judged intentional

<sup>4</sup> We note that in other experiments, these same factors – emotional salience, control, knowledge, and intent – both influence and may be influenced by moral judgments (Knobe, 2005, 2010). Experiment 2 does not conflict with prior evidence of this bi-directional influence. These results serve to show only that our scenarios describing accidental incest and accidental harm did not differ along these other dimensions.

<sup>5</sup> No other effects were significant, including all effects involving question order ( $p > 0.3$ ).

<sup>6</sup> As in our previous analysis and experiments, we also found main effects of domain ( $F(1, 152) = 7.9$ ,  $p = 0.006$ , partial  $\eta^2 = 0.05$ ) and intent ( $F(1, 152) = 33.3$ ,  $p < 0.001$ , partial  $\eta^2 = 0.18$ ), and the key interaction between domain and intent ( $F(1, 152) = 8.1$ ,  $p = 0.005$ , partial  $\eta^2 = 0.05$ ). Again, no other effects were significant, including all effects involving question order ( $p$ 's  $> 0.13$ ).



**Fig. 3.** Experiment 4: Attempts versus accidents. Moral judgments of the decision to act in the harm cases (light) and the decision to act in the incest cases (dark). Error bars represent standard error.

purity violations more morally wrong than accidental purity violations (intentional: 5.49; accidental: 4.50;  $t(81) = 2.4$ ,  $p = 0.02$ ) but not significantly more disgusting (intentional: 5.79; accidental: 5.27;  $t(81) = 1.5$ ,  $p = 0.2$ ). Participants also judged purity violations more disgusting than harm violations (purity: 5.52; harm: 4.40;  $t(158) = 3.5$ ,  $p < 0.001$ ) but not more morally wrong (purity: 4.96; harm: 4.36;  $t(158) = 1.6$ ,  $p = 0.10$ ).

The current participants judged purity violations significantly more disgusting than harm violations, consistent with evidence that feelings of disgust contribute to moral judgment especially in the domain of purity (Chapman et al., 2009; Eskine et al., 2011; Haidt et al., 1993; Helzer & Pizarro, 2011; Inbar, Pizarro, & Bloom, 2009; Inbar, Pizarro, Knobe et al., 2009; Schnall, Benton et al., 2008; Schnall, Haidt et al., 2008; Wheatley & Haidt, 2005; Zhong & Liljenquist, 2006). Importantly, though, the present results show that moral judgments and disgust ratings are not equivalent in the case of harm or purity violations. Participants did not simply interpret “how morally wrong” to mean “how disgusting” for our scenarios. Participants judged purity violations to be significantly more disgusting than harm violations, but they did not judge purity violations to be more morally wrong than harm violations. Also, the factor of intent influenced moral judgments and disgust ratings differently; participants perceived a significantly greater moral (versus disgust) difference between intentional and accidental purity violations.

## 6. Experiment 4: accidents and attempts

In Experiments 1–3, participants judged accidental harms to be less morally wrong, based on innocent intentions, than accidental purity violations. Is this effect specific to moral judgments of accidents, or does it reflect a general difference in the role of intent for moral judgments in these two domains? To address this question, Experiment 4 tested the complementary hypothesis that failed attempts to harm would be judged more morally wrong, based on guilty intentions, than failed attempts to violate purity norms (i.e., attempting but failing to commit incest).

In other words, Experiment 4 investigated whether moral judgments of harms would rely more on intent, and moral judgments of purity violations more on outcome, even for failed attempts.

### 6.1. Method

We collected data from 182 new participants (88 female, aged 19–70 yrs, mean = 35 yrs, standard deviation = 12 yrs), on Amazon Mechanical Turk. We eliminated 6 repeat participants, yielding a total of 177 participants. We presented four versions of the allergy (harm) and sibling incest (purity) scenarios (Fig. 1): (1) all-neutral (no violation), (2) accidental violation (neutral intent, bad outcome), (3) failed attempt (bad intent, neutral outcome), and (4) intentional violation (bad intent, bad outcome). Participants judged the moral wrongness of the *decision* to act in all conditions. Experiment 4 therefore used a 2 (domain: harm versus incest)  $\times$  2 (intent: neutral versus bad)  $\times$  2 (outcome: neutral versus bad) between-subjects design.

### 6.2. Results and discussion

Moral judgments were first analyzed in a 2 (domain: harm versus incest)  $\times$  2 (intent: neutral versus bad)  $\times$  2 (outcome: neutral versus bad) ANOVA (Fig. 3). We observed no main effect of domain ( $F(1, 175) = 0.07$ ,  $p = 0.80$ ). However, we observed main effects of intent (i.e., bad intentions were judged morally worse than neutral intentions,  $F(1, 175) = 78.7$ ,  $p < 0.001$ , partial  $\eta^2 = 0.32$ ) and outcome (i.e., bad outcomes were judged morally worse than neutral outcomes,  $F(1, 175) = 18.2$ ,  $p < 0.001$ , partial  $\eta^2 = 0.10$ ), as well as the key interaction between domain and intent ( $F(1, 175) = 9.34$ ,  $p = 0.003$ , partial  $\eta^2 = 0.05$ ) and the interaction between domain and outcome ( $F(1, 175) = 5.1$ ,  $p = 0.03$ , partial  $\eta^2 = 0.03$ ).<sup>7</sup> As in Experiments 1–3, the critical domain by intent interaction shows that the factor of intent (neutral versus bad)

<sup>7</sup> No other effects, including the three-way interaction, were significant.

matters differently for different moral domains (harm versus incest) – intent matters more for harm than for incest. Related to this interaction is the domain by outcome interaction: if participants are not relying on intent information as much in their moral judgments of incest, then they are relying more on outcome information. In other words, when intent is held constant, participants see a greater moral difference between bad and neutral outcomes for incest (e.g., intentional versus attempted incest; accidental incest versus all-neutral acts), versus harm. To clarify these interaction effects, we present below the separate analyses for harm and incest.

In two  $2$  (intent: neutral versus bad)  $\times$   $2$  (outcome: neutral versus bad) ANOVAs, we found that moral judgments are sensitive to intent in both domains (i.e., main effect of intent for harm:  $F(1, 98) = 107.7$ ,  $p < 0.001$ , partial  $\eta^2 = 0.53$ ; main effect for incest:  $F(1, 76) = 11.4$ ,  $p = 0.001$ , partial  $\eta^2 = 0.14$ ), but only moral judgments of incest are sensitive to outcome (i.e., main effect of outcome for incest:  $F(1, 76) = 65.8$ ,  $p < 0.001$ , partial  $\eta^2 = 0.16$ ). The intent by outcome interaction was not significant for harm or incest ( $p$ 's  $> 0.5$ ).

To further investigate the extent to which moral judgments of harm and incest are differently sensitive to the factors of intent and outcome, we conducted a series of independent-samples  $t$ -tests, comparing all four conditions, for harm and incest separately. Harm judgments showed a difference between all conditions that differed in intent only, but not outcome: intentional versus accidental harm (intentional harm: 6.50; accidental harm: 3.16;  $t(30) = 6.9$ ,  $p < 0.001$ ), all-neutral versus failed attempts to harm (all-neutral: 2.50; attempted harm: 6.00;  $t(26) = 6.9$ ,  $p < 0.001$ ). Harm judgments showed no difference between conditions that differed in outcome only, but not intent: all-neutral versus accidental harm ( $t(37) = 1.1$ ,  $p = 0.30$ ), intentional harm versus failed attempt to harm ( $t(58) = 1.5$ ,  $p = 0.15$ ). By contrast, incest judgments showed a difference between all conditions that differed in outcome only, but not intent: all-neutral versus accidental incest (all-neutral: 2.86; accidental incest: 4.40;  $t(40) = 2.1$ ,  $p = 0.04$ ), intentional versus failed attempt to commit incest (intentional incest: 6.40; attempted incest: 4.20;  $t(20) = 3.3$ ,  $p = 0.004$ ). Incest judgments showed no difference between two conditions that differed in intent only, but not outcome: all-neutral and failed attempt ( $t(35) = 1.7$ ,  $p = 0.10$ ).

Finally, we compared harm to incest for each of these four conditions: accidental, attempted, intentional, all-neutral. As in Experiments 1–3, accidental harms trended towards being judged less morally wrong than accidental incest ( $t(35) = 1.8$ ,  $p = 0.07$ ), based on innocent intentions. Also, as predicted, failed attempts to harm were judged morally worse than failed attempts to commit incest ( $t(17) = 3.6$ ,  $p = 0.01$ ), based on guilty intentions. There were no differences in judgments of harm and incest for the other conditions: intentional violations or all-neutral acts ( $p$ 's  $> 0.6$ ).

In sum, as predicted, failed attempts to harm were judged morally worse, based on false beliefs and guilty intentions, than failed attempts to commit incest. More generally, moral judgments of harm were more sensitive

to differences in intent than moral judgments of incest. Moral judgments of incest were more sensitive to differences in outcome.

## 7. Experiment 5: two kinds of failed attempts

In Experiments 1, 2, and 3, participants judged accidental incest more morally wrong than accidental harms; in Experiment 4, participants judged attempted incest less morally wrong than attempted harms. If intention plays a smaller role in moral judgments of incest, what is the other factor playing a larger role? One plausible hypothesis is that moral judgments of incest depend only on what *actually happened*: did incest actually occur or not? If so, then all cases in which incest does not actually happen (e.g., all failed attempts) should be judged morally equivalent. By contrast, we hypothesized that moral judgments of incest rely not just on what actually happened but also on what might have happened or what almost happened.

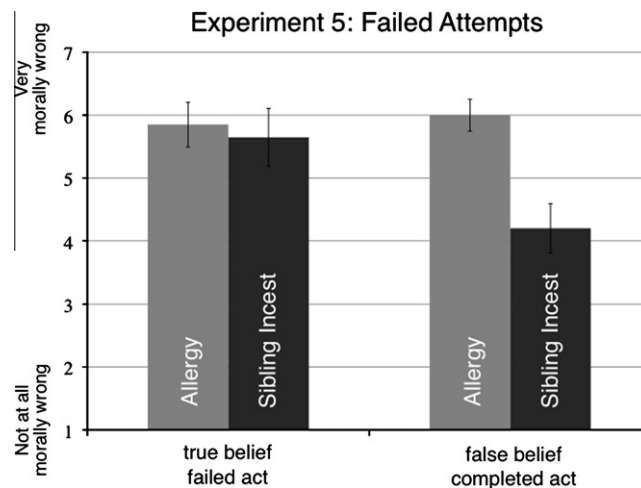
To test these two alternatives, we presented participants with scenarios depicting two kinds of failed attempts. The agent intended to commit incest (or harm) but failed either (1) because the completed act was not a moral violation after all, i.e. sleeping with a person *falsely* believed to be a sibling but who turns out to be unrelated (“false belief/completed act”) or (2) because the act could not be completed, i.e. almost sleeping with a *true* sibling but failing to do so because of an interruption, like a fire alarm (“true belief/failed act”).

In both cases, the agent *intends* to commit a moral violation but fails. As a consequence, we predicted that moral judgments of harm, which depend primarily on intentions, would be equivalent for these two kinds of failed attempts, characterized by the same bad intent. These two kinds of scenarios are also matched in terms of whether a moral violation actually occurred (e.g., neither case involved actual incest) but not in what *almost* happened (e.g., sex with a non-sibling, versus sex with a sibling). We predicted that participants would judge the decision to commit incest in these two cases differently and, specifically, more morally wrong if the act almost resulted in actual incest.

### 7.1. Method

We collected moral judgment data from 60 new participants (30 female, aged 18–70 yrs, mean = 33 yrs, standard deviation = 13 yrs), on Amazon Mechanical Turk. We eliminated 3 repeat participants, yielding a total of 57 participants. New participants were presented with failed attempts, involving true beliefs and interrupted acts (e.g., almost sleeping with an actual sibling but failing to do so because a fire alarm goes off), for both harm and incest (Fig. 1). A separate group of 40 participants also made judgments of how close the agent was to committing actual incest for the true belief versus false belief scenarios: How close were you to committing actual incest? (1 = not at all close, 7 = extremely close).





**Fig. 4.** Experiment 5: Failed attempts. Moral judgments of the decision to act in the Harm cases (light) and the decision to act in the incest cases (dark), for true belief – failed act condition (left) and false belief – completed act (right). Error bars represent standard error.

## 7.2. Results and discussion

First, as predicted, participants judged that the agent came significantly closer to committing actual incest in the case of true beliefs (mean: 6.0, standard deviation: 1.2) versus false beliefs (mean: 4.3, standard deviation: 2.6;  $t(27, \text{corrected d.f.}) = 2.7, p = 0.01$ ). That is, participants perceived failed attempts to commit incest, given *true* beliefs (e.g., almost sleeping with an actual sibling), to come closer to actual incest than failed attempts to commit incest, based on *false* beliefs (e.g., actually sleeping with a person falsely believed to be a sibling).

Second, we analyzed moral judgments in response to the two kinds of failed attempts – “true belief/failed act” versus “false belief/completed act” – using a 2 (true belief versus false belief)  $\times$  2 (harm versus incest) ANOVA (Fig. 4). We observed a marginal main effect of belief (i.e., true beliefs were judged worse than false beliefs,  $F(1, 111) = 3.7, p = 0.06$ , partial  $\eta^2 = 0.03$ ). We also observed a main effect of domain ( $F(1, 111) = 8.8, p = 0.004$ , partial  $\eta^2 = 0.08$ ), indicating that, overall, attempted harms were judged morally worse than attempted incest, as observed in Experiment 4, across both true and false beliefs. Finally, we observed the critical belief by domain interaction ( $F(1, 111) = 5.6, p = 0.02$ , partial  $\eta^2 = 0.05$ ). That is, for harm, both kinds of attempts were judged equally morally wrong (true belief: 5.85; false belief: 6.00;  $t(78) = 0.45, p = 0.66$ ), as predicted; it did not matter whether the agent truly or falsely believed that acting would cause harm.<sup>8</sup> By contrast, for incest, failed attempts to commit incest were judged morally worse in the case of a true belief/failed act than in the case of a false belief/completed act (true belief: 5.65; false belief: 4.20;  $t(21) = 2.1, p = 0.048$ ).

<sup>8</sup> As in Experiment 4, moral judgments of harm distinguished between the all-neutral act and attempted harm, based on a true belief ( $t(58) = 6.7, p < 0.001$ ), which differed in intent only, not outcome. Moral judgments of harm did not, however, distinguish between intentional and attempted harms, based on true beliefs, which differed in outcome only, and not intent, ( $t(58) = 1.6, p = 0.12$ ).

These findings show that failed attempts to harm are always judged very morally wrong, independent of whether the agent truly or falsely believes he or she will cause harm. What matters most for moral judgments of harm is the agent’s harmful intent. By contrast, failed attempts to commit incest are judged morally worse when the agent almost commits actual incest, than when the agent successfully completes an act that he or she only falsely believed to be incest. In other words, participants judge failed attempts that come closer to actual incest as morally worse.

## 8. General discussion

The current study reveals differences in the cognitive processes that support moral judgments in two distinct domains: harm versus purity. Intentions are assigned more weight for moral judgments of harm violations, like assault, compared to purity violations, like incest. In particular, we found that innocent intentions reduce blame for accidents (Experiments 1–3) and guilty intentions increase blame for failed attempts (Experiments 4–5) to a significantly greater extent for moral judgments of harm versus purity violations. Participants judge that it is morally wrong to intend harm, independent of the outcome, whereas moral judgments of purity violations depend more on what actually or almost occurs.

We observed the same pattern for moral scenarios presented in the second person (“You go to a party...”, Experiments 1A, 2–5) and in the third person (“Sam goes to a party...”, Experiment 1B), though we note that most of the present evidence was obtained using second-person scenarios. The specific behavioral patterns found in Experiments 2–5 should therefore be replicated using third-person scenarios. While prior work has investigated moral judgments of one’s own intentional harms versus another person’s intentional harms (Kedia et al., 2008), future work should directly test the impact of perspective (e.g., self versus other) on the role of *intent* in moral judgments.

We consider here why accidental purity violations, like accidental incest, might be perceived more morally wrong than accidental harms. Our results rule out explanations based on a number of possible confounds. For example, participants might have assumed (contrary to the information provided in the stimuli) that one could never truly commit accidental incest, because the sibling or parent would inevitably be recognized. Relatedly, participants might have seen agents committing “accidental” incest as doing so knowingly and intentionally. Contrary to these concerns, however, participants explicitly judged that the agents in our scenarios *knew* what they were doing, had *control* over their actions, and acted *intentionally*, to the same extent for harm and purity violations.

What is the role of participants’ emotions in generating these moral judgments? At the simplest level, emotional responses could also be a confound in the experiment: if participants are generally more upset by the idea of incest, as opposed to assault, for example, they might be less capable of making sophisticated cognitive inferences, which are required for intent-based moral judgments. Once again, the empirical data do not support this concern: participants reported being equally emotionally upset by the scenarios describing incest and harm. More specifically, participants found the act of accidentally poisoning a friend to be equally emotionally upsetting as accidentally sleeping with a long-lost sibling. This result is consistent with prior behavioral and neural evidence indicating robust emotional responses to purely harmful (but not canonically disgusting) actions (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Kedia et al., 2008; Strohminger, Lewis, & Meyer, 2011; Valdesolo & DeSteno, 2006; Young et al., 2010).

A more complex question concerns the role of a specific emotion: disgust. Overall, people (including our participants) find purity violations (e.g., incest) to be more disgusting than harmful acts (Haidt et al., 1993; Royzman et al., 2009). Indeed, recent research suggests that the specific emotion of disgust plays a key role in moral judgment (Hsu, Anen, & Quartz, 2008; Zhong & Liljenquist, 2006), and especially in the domain of purity (Chapman et al., 2009; Haidt et al., 1993; Helzer & Pizarro, 2011; Inbar, Pizarro, & Bloom, 2009; Inbar, Pizarro, Knobe et al., 2009; Schnall, Benton et al., 2008; Schnall, Haidt et al., 2008; Wheatley & Haidt, 2005). For example, in one study, individual differences in disgust responses correlated with moral judgments: political conservatives appeared to be more “disgust sensitive” (Inbar, Pizarro, & Bloom, 2009), and this measure correlated with moral attitudes towards homosexuality (Inbar, Pizarro, Knobe et al., 2009). Another study hypnotized participants to experience disgust at an arbitrary word (Wheatley & Haidt, 2005). When this word was used to describe actions (e.g., second cousins who had a sexual relationship), participants delivered harsher moral judgments. In a third study that focused on participants’ own behavior as opposed to their judgment of others, participants were more likely to engage in immoral behaviors after cleaning their hands, and, reciprocally, participants were more likely to engage in physical cleansing after behaving immorally (Zhong & Liljenquist, 2006). In all, these findings suggest that disgust is a key influence on judgments of moral purity.

Could moral judgments of purity violations be less sensitive to information about the agent’s intent because of the influence of disgust? This suggestion has a strong and a weak interpretation. On the strong interpretation, participants’ moral judgments of purity violations are simply identical to their expressions of disgust. That is, participants use the moral judgment question as an opportunity to express their feelings of disgust.<sup>9</sup> We provide evidence against this strong interpretation: participants’ ratings showed significantly different patterns, across our scenarios, in response to the questions “how morally wrong” versus “how disgusting” for each action. On a weaker interpretation, moral judgments of purity violations are crucially influenced by feelings of disgust, as prior evidence suggests. Feelings of disgust are not affected by intentionality, so accidental incest is judged just as disgusting as intentional incest. Thus, it may be the key feeling of disgust that focuses participants’ attention on the act itself, and therefore partly dampens their consideration of the intention.

Nevertheless, in the current study, participants’ moral judgments of purity violations appear to reflect a sophisticated calculation, an integration of multiple cognitive inputs, rather than a simple gut reaction. In particular, participants were able to differentiate between two scenarios containing identical intentions (to commit incest) and identical outcomes (no incest was committed) in Experiment 5. The decision to commit incest was judged morally worse when the agent *truly* believed he or she was about to commit incest, relative to a failed attempt based on a *false* belief.<sup>10</sup> That is, nearly sleeping with a true sibling was seen as morally worse than successfully sleeping with a stranger falsely believed to be a sibling. These judgments indicate participants’ consideration of the counterfactual: how close did the agent come to committing *actual* incest? Participants are not simply reacting emotionally to the idea of incest; they are making moral judgments by combining multiple relevant factors (Lieberman, Tooby, & Cosmides, 2003, 2007). Among these factors, intention is considered but is given less weight than in moral judgments about harm.

These arguments suggest a broader question concerning the content of morality. What does it mean to deliver a specifically *moral* judgment? One interpretation of the current results is that moral judgment is not a unitary feature of human psychology (Monin et al., 2007; Nichols, 2002; Sinnott-Armstrong, 2008). Instead, “moral wrongness” could take on qualitatively different meaning depending on the context, just as the polysemous term “bank” refers sometimes to a place where money is kept and sometimes to a part of a river. We agree that moral judgments of different kinds of actions depend differently on properties like intent, just as the current results show. However, we maintain that people can still make genuinely *moral* judgments across domains. Though the current

<sup>9</sup> Another possibility is that the participants viewed the disgust rating question as an opportunity to express moral disapproval (Danovitch & Bloom, 2009).

<sup>10</sup> By contrast, an agent’s decision to cause harm was judged equally morally wrong, whether the agent believed, *truly* or *falsely*, that he or she would cause harm.

state of moral psychology provides no definitive answers, we suggest that some cognitive content or computation may apply to all moral judgments, and specifically to moral judgments (DeScioli & Kurzban, 2009; Gray & Wegner, 2009; Mikhail, 2007; Royzman et al., 2009; Turiel, 1983; Walker & Hennig, 2004). For example, moral judgments, including of purity violations, may share certain features, compared to other kinds of social violations (e.g., violating conventions of dress): people believe that acts that are morally wrong are wrong at any time, in any place, and even if an authority figure or society were to specifically condone the act (Turiel, 1983). Recent research indicates some exceptions to these criteria (Kelly, Stich, Haley, Eng, & Fessler, 2007; Nichols, 2002). Thus, investigating which cognitive or neural processes are common to all kinds of moral judgment (and possibly unique to moral judgment) will be an important challenge for the future of moral psychology (Royzman et al., 2009; Walker & Hennig, 2004; Young & Dungan, in press).

In the current study, we focused on the domains of harm and purity because they represent well-studied domains within moral psychology, but future work should also investigate the role of intent for other moral domains (Suhler & Churchland, 2011). Intent may play different roles in the domains of fairness, in-group loyalty, and respect for authority (Graham & Haidt, 2010; Haidt, 2007; Rozin et al., 1999), as well as for norms that may be inextricably linked to intent, such as norms against dishonesty (Greene & Paxton, 2009; Mazar, Amir, & Ariely, 2008). In fact, understanding how moral judgments may be differently governed by cognitive rules (e.g., intentionality) may lead moral psychologists to more precisely define domain boundaries (Wright & Baril, 2011).

The existence of distinct cognitive signatures of different moral domains may provide clues concerning the functional roles of moral judgment, in individuals and societies. Norms concerning harm and purity may promote social cohesion through different means (DeScioli & Kurzban, 2009; Graham & Haidt, 2010; Haidt, 2001; Rai & Fiske, 2011; Rozin et al., 1999; Shweder & Haidt, 1993). Moral norms may help us identify social partners who (a) behave well towards us and (b) behave similarly to us – both key ingredients for successful social groups. One possibility is that harm norms primarily guarantee that people will behave well towards us in the future, whereas purity norms serve to increase behavioral homogeneity. These speculations could shed light on the current data. In the case of incest avoidance and food taboos, the act itself, rather than the underlying intent, may be a signal of group membership: do you eat what we eat, and behave as we do? Eating the wrong food or affiliating with the wrong people may indicate the threat of moral heterogeneity (Atran & Henrich, 2010; Sosis & Bressler, 2003). Purity norms may therefore attach little weight to individuals' intentions. Indeed, in some cultures, even the *victim* of a purity violation is punished, as in honor killings of rape victims (Rai & Fiske, 2011). By contrast, for harm norms to regulate future behavior, intent is critical. Especially in the case of accidents, only knowing people's intentions can lead to reliable predictions of their future behavior (Waytz et al., 2010), and accurate identification of true friends and foes.

On an alternative account, rules against incest and taboo foods may have developed as a means for individuals to protect *themselves*, for their own good, from possible contamination (Chapman et al., 2009; Rozin et al., 1999). When we worry about negatively impacting ourselves, we may care less about whether the impact is accidental or intentional; the key is to avoid the contamination. Purity violations may therefore be evaluated on the basis of the actual or anticipated outcome, not intent. By contrast, harm norms may have developed to guide people to behave well towards *others* and to limit their negative impact on one another. Intent often functions in such cases to identify appropriate social partners: information about intent supports not only explanations and evaluations of past behavior, but also reliable predictions of future behavior (e.g., who will or will not hurt me in the future), as noted above. Indeed, reasoning about intent (and other mental states) appears to be particularly salient in interpersonal contexts (e.g., I hurt you; you hurt me) (Baumeister et al., 1994; Kedia et al., 2008). These distinct functional origins may therefore help explain the current results concerning the role of intent across distinct moral domains.

Of course, these theoretical accounts go far beyond the current behavioral data. Future research on the psychology of morality must give careful consideration to the roles of different moral norms across different cultures. Already, a few studies have reported that cultures may differ in the moral weight they assign to an individual agent's beliefs and intentions (Cohen & Rozin, 2001; Hamilton & Sanders, 1983; Mazar & Aggarwal, in press). In one study, Jews weighed outcomes more than mental states, endorsing actions performed out of obligation rather than good will, such as resentfully caring for one's parents; Christians, by contrast, weighed mental states more, rejecting such actions where "one's heart isn't in it" as hypocritical (Cohen & Rozin, 2001). In another study, Japanese participants, compared to American participants, assigned less weight to personal intentions (Hamilton & Sanders, 1983). These cultural differences may relate to the domain differences we have reported here: if a group or society places heavier emphasis on the avoidance of purity violations (Graham & Haidt, 2010; Haidt & Hersh, 2001; Inbar, Pizarro, & Bloom, 2009; Inbar, Pizarro, Knobe et al., 2009), for which intent is less relevant, then its moral norms as a whole may assign less weight to intent. Indeed, different individuals, across different cultures, may disagree about whether a behavior is right or wrong, why it is wrong, and even whether the primary issue is moral or prudential. While psychological research may not solve the problem of moral disagreement, it may help understand its origins.

## Acknowledgements

We thank James Dungan, Alek Chakroff, Jonathan Haidt, Joe Henrich, Clark Barrett, Kurt Gray, Adam Waytz, Laura Schulz, Walter Sinnott-Armstrong, and Fiery Cushman, and the members of the *Culture and the Mind Project*, funded by the Arts and Humanities Research Council (Project Director: Stephen Laurence, University of Sheffield), for useful discussions. The authors were supported by the

Simons Foundation, the John Merck Scholars Program, and the Packard Foundation.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.cognition.2011.04.005](https://doi.org/10.1016/j.cognition.2011.04.005).

## References

- Atran, S., & Henrich, J. (2010). The evolution of religion: HOW cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biological Theory*, 5(1), 18–30.
- Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, 103, 37–49.
- Baumeister, R. F., Stillwell, A., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115, 243–267.
- Blair, K., Marsh, A., Finger, E., Blair, S., & Luo, J. (2006). Neuro-cognitive systems involved in morality. *Philosophical Explorations*, 9(1), 13–27.
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, 18(5), 803–817.
- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, 323(5918), 1222–1226.
- Cohen, A. B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, 81(4), 697–710.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analysis in moral judgment. *Cognition*, 108(2), 353–380.
- Danovitch, J., & Bloom, P. (2009). Children's extension of disgust to physical and moral events. *Emotion*, 9(1), 107–112.
- DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, 112(2), 281–299.
- Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological Science*, 22(3), 295–299.
- Graham, J., & Haidt, J. (2010). Beyond beliefs: Religion binds individuals into moral communities. *Personality and Social Psychology Review*, 14, 140–150.
- 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., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, 96(3), 505–520.
- Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 106(30), 12506–12511.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31(1), 191–221.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65(4), 613–628.
- Hamilton, V., & Sanders, J. (1983). Universals in judging wrongdoing: Japanese and Americans compared. *American Sociological Review*, 48, 199–211.
- Hart, H. L. A. (1968). *Punishment and responsibility*. Oxford: Oxford University Press.
- Helzer, E. G., & Pizarro, D. A. (2011). Dirty liberals!: Reminders of physical cleanliness influence moral and political attitudes. *Psychological Science*.
- Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, 320(5879), 1092–1095.
- Inbar, Y., Pizarro, D., & Bloom, P. (2009a). Conservatives are more easily disgusted than liberals. *Cognition and Emotion*, 23, 714–725.
- Inbar, Y., Pizarro, D. A., Knobe, J., & Bloom, P. (2009b). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion*, 9(3), 435–439.
- Kedia, G., Berthoz, S., Wessa, M., Hilton, D., & Martinot, J. L. (2008). An agent harms a victim: A functional magnetic resonance imaging study on specific moral emotions. *Journal of Cognitive Neuroscience*, 20(10), 1788–1798.
- Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect, and the moral/conventional distinction. *Mind and Language*, 22(2), 117–131.
- Killen, M., Lynn Mulvey, K., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, 119(2), 197–215.
- Knobe, J. (2005). Theory of mind and moral cognition: Exploring the connections. *Trends in Cognitive Sciences*, 9, 357–359.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, 315–329.
- Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 270(1517), 819–826.
- Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, 445(7129), 727–731.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental and Social Psychology*, 33, 101–121.
- Mazar, N., & Aggarwal, P. (in press). Can collectivism promote bribery? *Psychological Science*.
- Mazar, N., Amir, O., & Arieli, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *American Marketing Association*, XLV(633–644).
- Mikhail, J. M. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, 11(4), 143–152.
- Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignacio, F. A., Bramati, I. E., Caparelli-Daquer, E. M., et al. (2005). The moral affiliations of disgust. *Journal of Cognitive Behavioral Neurology*, 18(1), 68–78.
- Monin, B., Pizarro, D., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology*, 11, 99–111.
- Nichols, S. (2002). Norms with feeling: Toward a psychological account of moral judgment. *Cognition*, 84, 221–236.
- Piaget, J. (1965). *The moral judgment of the child*. New York: Free Press.
- Rai, T., & Fiske, A. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality and proportionality. *Psychological Review*, 118(1), 57–75.
- Rozyman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Toward a content-specific account of the moral-conventional distinction. *Cognition*, 112(1), 159–174.
- 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.
- Schaich Borg, J., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience*, 20(9), 1529–1546.
- Schnall, S., Benton, J., & Harvey, S. (2008a). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, 19(12), 1219–1222.
- Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008b). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096–1109.
- Shweder, R., & Haidt, J. (1993). The future of moral psychology: Truth, intuition, and the pluralist way. *Psychological Science*, 4, 360–365.
- Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, 41(4), 653–662.
- Sinnott-Armstrong, W. (2008). Is moral phenomenology unified? *Phenomenology and the Cognitive Sciences*, 7(1), 85–97.
- Sosis, R., & Bressler, E. (2003). Cooperation and commune longevity: A test of the costly signaling theory of religion. *Cross-Cultural Research*, 37, 211–239.
- Strohlinger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition*, 119(2), 295–300.
- Suhler, C. L., & Churchland, P. (2011). Can innate, modular “foundations” explain morality? Challenges for Haidt's moral foundations theory. *Journal of Cognitive Neuroscience*.

- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476.
- Walker, L., & Hennig, K. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology*, *86*(4), 629–647.
- Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010a). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388.
- Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J. H., & Cacioppo, J. T. (2010b). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, *99*(3), 410–435.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*(10), 780–784.
- Wright, J., & Baril, G. (2011). The role of cognitive resources in determining our moral intuitions: Are we all liberals at heart? *Journal of Experimental Social Psychology*.
- Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, *65*, 845–851.
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240.
- Young, L., & Dungan, J. (in press). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*.
- Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*(5792), 1451–1452.