Article

Social Psychological and
Personality Science
2015, Vol. 6(8) 859-868
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1948550615592241
spps.sagepub.com
SAGE

Impure or Just Weird? Scenario Sampling Bias Raises Questions About the Foundation of Morality

Kurt Gray[1] and Jonathan E. Keeney[2]

Abstract

Moral psychologists have used scenarios of abuse and murder to operationalize harm and chicken-masturbation and dog-eating to operationalize impurity. These scenarios reveal different patterns of moral judgment across harm and purity, ostensibly supporting distinct moral mechanisms, modules, or "foundations." However, these different patterns may stem not from differences in moral content per se but instead from biased sampling that confounds content with weirdness and severity. Supporting this hypothesis, frequently used impurity scenarios are *weirder* and *less severe* than both harm scenarios (Study 1) and participant-generated impurity scenarios (Study 2). Weirdness and severity—not content—also appear to drive differences between act and character evaluations (Study 3). Also problematic for modular accounts are extremely high correlations between harm and impurity (*rs* > .86), and findings that harm scenarios assess impurity better than researcher-devised impurity scenarios. Overall, patterns of moral judgment previously ascribed to distinct moral mechanisms may reflect domain-general moral cognition.

Keywords

morality, values, moral foundations, dyadic morality, purity

In 1936, the United States was in the midst of a closely contested presidential election between Alf Landon and Franklin Roosevelt. Using automobile and telephone records, the magazine *Literary Digest* confidently predicted a Landon victory. Unfortunately, car and phone owners were unrepresentative of the general population because of their relative affluence. Roosevelt easily won the election, and within 2 years *Literary Digest* was defunct. The lesson here is clear—unrepresentative samples can invalidate conclusions, a problem called *sampling bias*. In moral psychology, sampling bias is problematic not only when selecting participants (e.g., recruiting only college freshmen; Henrich, Heine, & Norenzayan, 2010) but also when selecting stimuli. In this article, we investigate sampling bias—that is, the presence of confounds—within popular moral scenarios that are often used to provide support for modular moral cognition.

The Structure of Moral Cognition

Is moral judgment a product of one process or many? Historically, moral judgment was thought to revolve only around direct physical and emotional harm, but anthropological reports suggest a diversity of moral content across cultures (Haidt, Koller, & Dias, 1993; Rai & Fiske, 2011). Some researchers explain these cultural differences by positing a whole number (typically between three and six) of *domain-specific* "cognitive modules," defined as "little switches in the brains of all animals" that are "triggered" by specific moral "inputs" (Haidt, 2012, p. 123). These modular theories—such as Moral Foundations Theory (MFT; Graham et al., 2012)—argue for differences in judgment based upon moral content per se, such that judgments about harm (inflicting physical and emotional suffering) involve fundamentally "distinct cognitive computations" (Young & Saxe, 2011, p. 203) than those regarding purity (violations of spirit or body; Graham et al., 2012). In contrast to modular accounts, *domain-general* accounts deny the existence of distinct moral modules, emphasizing instead common affective and conceptual considerations—that is, dimensions—that overarch all moral content (for a review see Cameron, Lindquist, & Gray, 2015).

Ostensible support for moral modules comes from studies revealing different patterns of judgment for harm and purity violations. For example, intention appears to be more important

[1] Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2] Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Corresponding Author:
Kurt Gray, Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
Email: kurtgray@unc.edu

for judgments concerning harm (e.g., murder) versus impurity (e.g., drinking urine; Young & Saxe, 2011), and character evaluations appear more related to impurity (e.g., dog-eating) than to harm (e.g., meat-stealing; Uhlmann & Zhu, 2014).

These apparent examples of modularity are intuitively compelling, but the researcher-generated scenarios they rely upon may suffer from sampling bias. Just as *Literary Digest* sampled unrepresentative people, these studies may have sampled unrepresentative impurity scenarios, introducing central confounds into their studies. Drinking urine and eating dog may be both less typical—that is, weirder—and less morally severe than harm violations such as murder. The continuous dimensions of severity and weirdness could give rise to different patterns of judgment, but through domain-general processes rather than distinct content-based mechanisms.

Severity—the moral extremity of an act—is perhaps the most important feature of a moral violation. By definition, more severe acts are more immoral; that is, they are better examples of the category "immorality" and are therefore more likely to engage moral cognition (Murphy, 2004). If researcher-generated impurity scenarios are especially mild, they may reveal different patterns of moral judgment—not because they are *differently* immoral but because they are merely *less* immoral. For example, different patterns for the role of intention in urine-drinking and murder could simply reflect the lesser severity of urine-drinking (Young & Saxe, 2011). Intention could be equally important to all moral content but dependent upon baseline immorality. If one act is half as immoral as another, we might expect intention to matter only half as much.

Weirdness—the extent to which an act is weird, bizarre, or unusual—may also affect the cognitive processing of moral judgments. Weird events are not only generally rare but are counternormative in ways beyond immorality, making them difficult to explain. One can more readily imagine motives for murder than for urine-soaked performance art or plastic surgery tails. According to attribution theory (Jones & Davis, 1965), inexplicable behaviors are more likely to be attributed to the actor's disposition (i.e., character; Pizarro & Tannenbaum, 2011) than to the situational factors. If impurity is confounded with weirdness, it would give the appearance that impurity influences judgments of character more than harm. Weirdness may also help account for the oft-discussed link between disgust and impurity (Gutierrez & Giner-Sorolla, 2007; Pizarro, Inbar, & Helion, 2011; but see Chapman & Anderson, 2014), without referencing distinct moral mechanisms. Just as novel and bizarre foods are often disgusting (Pliner & Pelchat, 1991) so too might be novel and bizarre moral violations.

The Current Research

First, we investigate whether the most popular scenarios assessing harm and impurity—those of MFT—confound moral content with severity and weirdness (Study 1). Second, we compare researcher-generated MFT scenarios with participant-generated scenarios to test whether these more naturalistic purity scenarios are relatively more severe and less weird (Study 2). Finally, we use these naturalistic scenarios to examine whether the dimensions of severity and weirdness can better explain the apparent link between impurity and character evaluations (Study 3).

We investigate impurity scenarios because of their pervasive role in moral psychology. At last count, researcher-generated scenarios have been used to operationalize impurity—and to buttress claims of modular morality—in 53 studies from 29 different articles, with 4,351 total citations (see Supplementary Table S1 in the Supplementary Materials). Twenty-six of those articles have used scenarios developed by MFT researchers (e.g., Haidt et al., 1993). Importantly, dismissals of domain-general moral cognition rely heavily on these self-report scenarios (p. 104, Graham et al., 2012). Such dismissals may be premature if moral content is confounded with severity and weirdness, especially if these two broad—and more parsimonious—dimensions can explain different patterns of judgment.

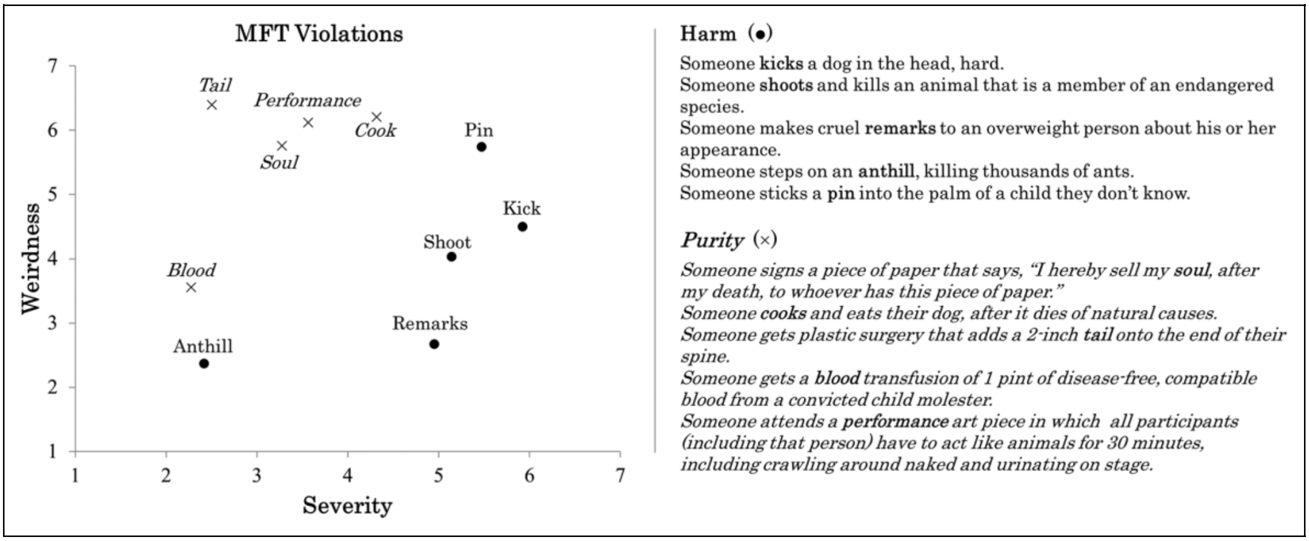
Study 1: MFT Scenarios

In this study, we examine whether the commonly used MFT harm and impurity scenarios (Graham, Haidt, & Nosek, 2009) confound moral content with severity and weirdness. We also examined the implicit—but surprisingly untested—modular claim that harm and purity violations activate distinct moral concerns: Harm violations should activate harm concerns but not impurity concerns, and vice versa (Graham et al., 2012). Thus, this study serves as a manipulation check for MFT scenarios, measuring whether they actually represent their content labels.

Method

Ninety-nine participants were recruited through Amazon's mTurk. Twenty-one failed to finish, and nine failed attention checks, leaving 69 (42% male, $M_{\text{age}} = 34$, 58% liberal).

All participants evaluated 10 MFT scenarios—5 harm violations and 5 purity violations (Figure 1; Supplementary Table S2, Supplementary Materials)—in random order. For each scenario, participants rated moral wrongness ("How morally wrong is this act?"), severity ("How severe is this act?") weirdness ("How atypical [i.e., weird, strange, unusual] is this act?"), harm ("How harmful [i.e., involving physical and/or emotional suffering] is this act?), and impurity ("How impure [i.e., involving sinfulness, indecency, dirtiness] is this act?") using 7-point scales from (1) *not at all wrong/severe/atypical/harmful/impure* to (7) *very wrong/severe/atypical/harmful/impure*. Definitions of harm and impurity were drawn directly from MFT research (Graham et al., 2009; Haidt & Graham, 2007; Haidt & Joseph, 2004, 2007). After evaluating scenarios, participants reported political orientation, using a 7-point scale from (1) *Strongly liberal* to (7) *Strongly conservative*.



Harm (●)

Someone **kicks** a dog in the head, hard.
Someone **shoots** and kills an animal that is a member of an endangered species.
Someone makes cruel **remarks** to an overweight person about his or her appearance.
Someone steps on an **anthill**, killing thousands of ants.
Someone sticks a **pin** into the palm of a child they don't know.

Purity (×)

*Someone signs a piece of paper that says, "I hereby sell my **soul**, after my death, to whoever has this piece of paper."*
*Someone **cooks** and eats their dog, after it dies of natural causes.*
*Someone gets plastic surgery that adds a 2-inch **tail** onto the end of their spine.*
*Someone gets a **blood** transfusion of 1 pint of disease-free, compatible blood from a convicted child molester.*
*Someone attends a **performance** art piece in which all participants (including that person) have to act like animals for 30 minutes, including crawling around naked and urinating on stage.*

Figure 1. Commonly used Moral Foundations Theory (MFT) scenarios (Graham et al., 2009), by severity and weirdness (Study 1).

Results

Scenario means and zero-order correlations are given in Supplementary Tables S3 and S4 in Supplementary Materials. To accommodate the hierarchical, nonindependent structure of these data, we analyzed multilevel, random intercept models (MLM models; McCulloch, Searle, & Neuhaus, 1998) with scenarios (Level 1) nested within participants (Level 2). In the first step, fixed main effects of moral content (coded as 0 = *harm*, 1 = *purity*; Level 1) and political orientation (standardized; Level 2) were estimated, and in the second step, a Content × Politics Cross-Level interaction term was included. To conserve text, we report only the effects of moral content only at Step 1 (See Supplementary Materials for all analyses involving politics).

As hypothesized, compared to MFT harm scenarios, MFT impurity scenarios were seen as less severe, $b = -1.60$, β(standardized) $= -.360$, $t(620) = -10.76$, $p < .001$, and weirder, $b = 1.74$, $\beta = .412$, $t(620) = 12.27$, $p < .001$. MFT harm scenarios were also perceived to be relatively more harmful, $b = -2.17$, $\beta = -.483$, $t(620) = -15.25$, $p < .001$ and—surprisingly—*more impure* than MFT impurity scenarios, $b = -1.12$, $\beta = -.256$, $t(620) = -7.36$, $p < .001$.

Discussion

Commonly used researcher-generated MFT scenarios confound moral content with severity and weirdness: Purity violations were substantially less severe and weirder than harm violations. Strikingly, MFT impurity scenarios appeared to fail their own manipulation check; harm scenarios involved *greater* impurity than custom-designed impurity scenarios, perhaps because harm is the essence of immorality (Gray, Young, & Waytz, 2012), and "impure"—that is, "sinful" or "indecent"—is synonymous with "morally wrong" (Oxford English Dictionary, n.d.).

Also noteworthy is that harm and impurity ratings were highly correlated in these scenarios, $r(8) = .89$, $p < .001$, casting doubt on the distinctness of these moral concerns. Of course, participants may be limited in their ability to accurately report their moral intuitions, but modular accounts of morality—with their anthropological roots—have long privileged the intuitions of participants over those of researchers (Haidt et al., 1993). In this case, such self-reports argue against distinct moral concerns, as do recent implicit studies, which find that impurity violations automatically activate harm (Gray, Schein, & Ward, 2014).

Study 2: Naturalistic Moral Violations

Study 1 revealed that MFT impurity scenarios are weirder and less severe than MFT harm scenarios. As harm is prototypical within moral cognition (Gray & Schein, 2012), it makes sense that harm should be somewhat more severe and typical than impurity (i.e., more representative of "immorality"). Nevertheless, researcher-generated MFT scenarios may exaggerate these differences and therefore fail to accurately represent the moral intuitions of laypeople (Inbar & Lammers, 2012). To test this hypothesis, we compared naturalistic participant-generated scenarios to MFT scenarios.

Scenario Generation

Two hundred ninety participants were recruited through mTurk. Fourteen failed to finish, leaving 276 (48% male, $M_{\text{age}} = 32$, 53% liberal). Participants were randomly assigned to volunteer either three harmful violations ("harmful, hurtful, damaging, or causing physical or emotional suffering") or three impure violations ("sinful, dirty, degrading, lustful, or indecent")—definitions again taken directly from modular accounts (Graham et al., 2009; Haidt & Joseph, 2004, 2007).

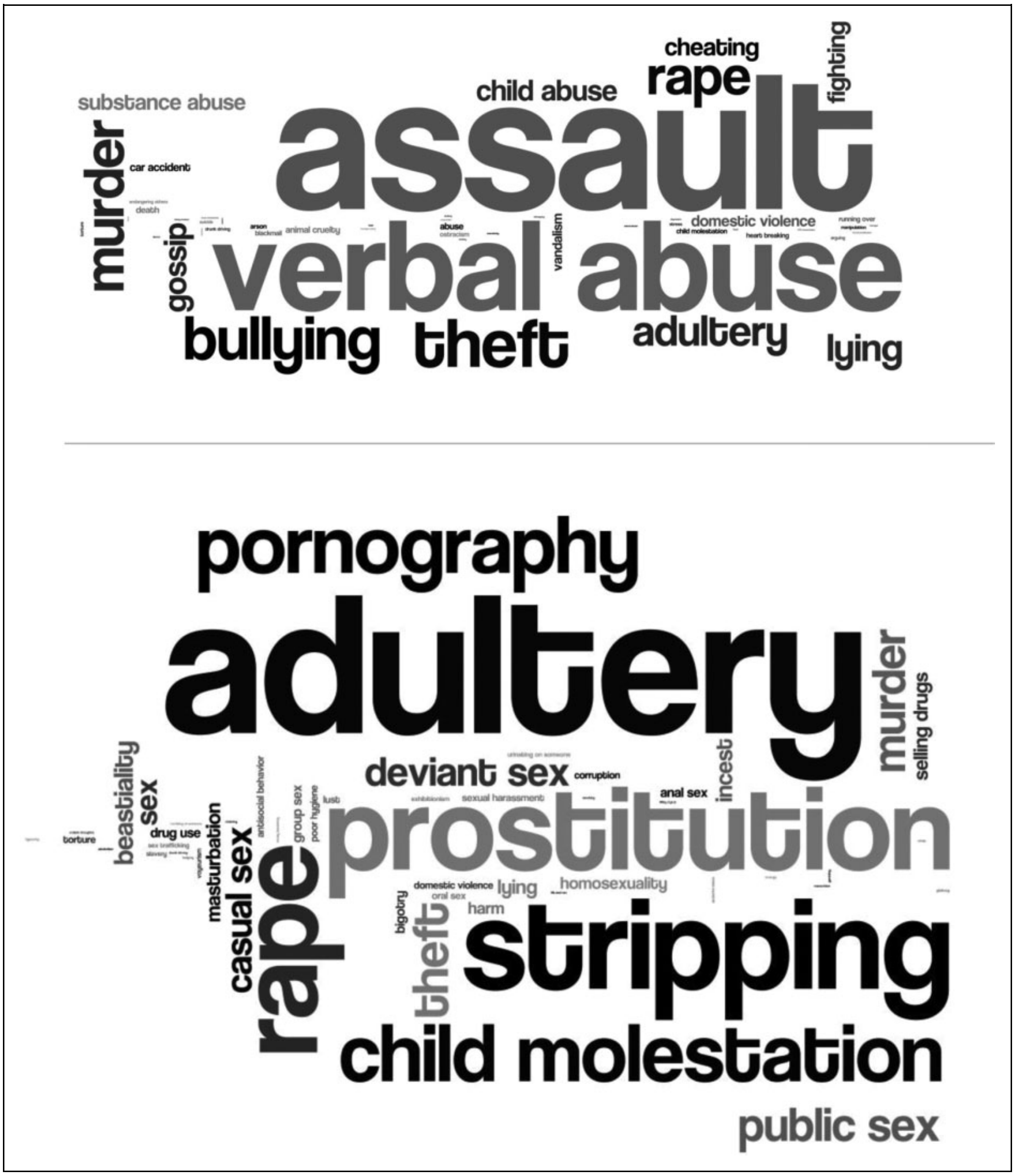


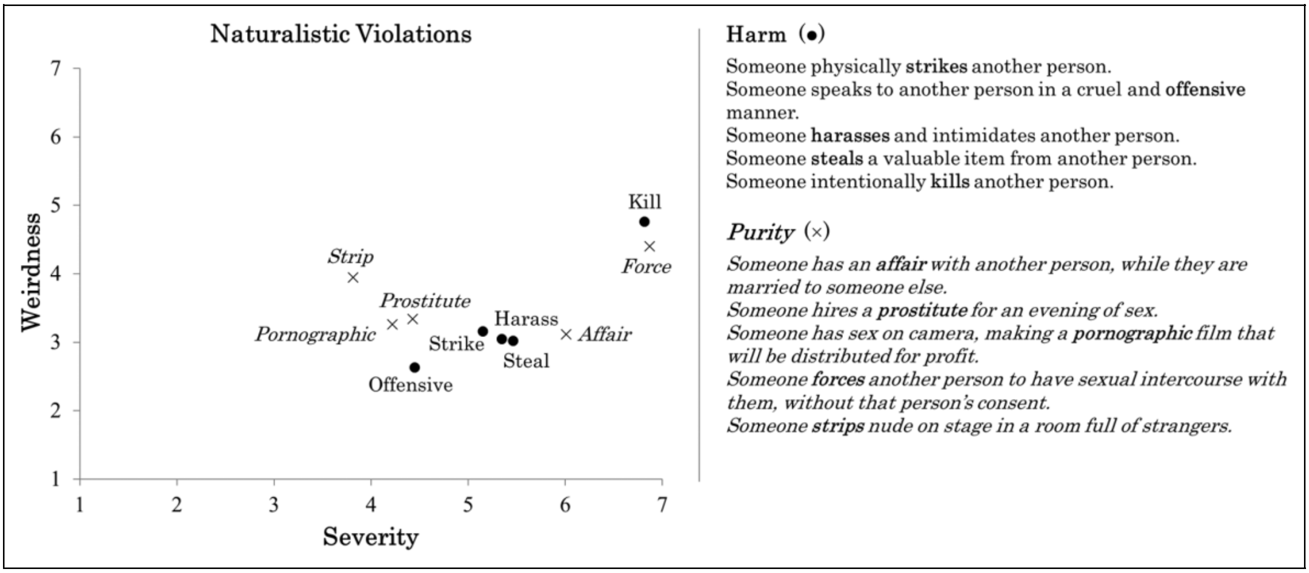
Figure 2. Word map of harm (top) and purity (bottom) violations generated by participants (Study 2). Text size indicates frequency.

Responses that clearly referred to the same violation were recoded and combined (e.g., ‘‘rape’’ and ‘‘forced sex’’); however, the vast majority of violations were unedited. See Figure 2 for word maps of participant responses (See Supplementary Table S5 in Supplemental Materials for frequencies). Notably, no participant generated dog-eating, chicken-masturbation, urine-drinking, or soul-selling as purity violations.

Based on the most popular violations volunteered by participants, we generated five new harm scenarios and five new impurity scenarios (Figure 3; Table S6 in Supplementary Materials). Although some of these violations (e.g., stripping, pornography, and prostitution) are ‘‘pure’’ impurity, other violations involve other moral content such as harm (e.g., rape) and loyalty (e.g., adultery), and so some might argue that they are not valid representations of purity per se. Nevertheless, our participant-centric approach is consistent with the anthropological roots of modular accounts (Shweder & Haidt, 1994). These ‘‘mixed’’ violations demonstrate that the moral intuitions of laypeople do not respect the firm content boundaries drawn by modular accounts of those intuitions (Haidt & Joseph, 2005).[1]

Scenario Comparison

One hundred and thirty-two participants were recruited through mTurk. Twenty-two failed to finish and three failed attention checks, leaving 107 (50% male, $M_{\text{age}} = 38$, 54% liberal). Participants provided ratings of severity, weirdness, harm, and impurity for all 20 violations—10 naturalistic and 10 MFT—and rated their political orientation, all as in Study 1. We predicted that the naturalistic impurity scenarios would be



Harm (●)

Someone physically **strikes** another person.
Someone speaks to another person in a cruel and **offensive** manner.
Someone **harasses** and intimidates another person.
Someone **steals** a valuable item from another person.
Someone intentionally **kills** another person.

Purity **(×)**

Someone has an ***affair*** *with another person, while they are married to someone else.*
Someone hires a ***prostitute*** *for an evening of sex.*
Someone has sex on camera, making a ***pornographic*** *film that will be distributed for profit.*
Someone ***forces*** *another person to have sexual intercourse with them, without that person's consent.*
Someone ***strips*** *nude on stage in a room full of strangers.*

Figure 3. Naturalistic scenarios, by severity and weirdness (Study 2).

evaluated as less weird and more severe than the MFT impurity scenarios, suggesting some sampling bias in MFT scenarios.

Results

Zero-order correlations and scenario means are provided in Supplementary Tables S7 and S8. See Figure 3 for a plot of naturalistic scenarios. For analysis, we used the same MLM models as in Study 1, with scenarios (Level 1) nested within participants (Level 2). In the first step, we estimated fixed main effects of moral content (coded as 0 = *harm*, 1 = *purity*; Level 1), scenario source (coded as 0 = *naturalistic*, 1 = *MFT*; Level 1), and political orientation (standardized; Level 2). In the second step, we entered interaction terms for Content × Source, Content × Politics, and Source × Politics. In the third step, we entered the three-way Content × Source × Politics interaction. To conserve text, we report effects of moral content and scenario source at Step 1, and the Content × Source interaction at Step 2 (i.e., the nonconditional interaction). The Supplementary Materials detail all politics main effects and interactions (note that the three-way interaction was not significant for any of the following analyses).

Severity

Across all 20 scenarios, impurity scenarios were less severe than harm scenarios, $b = -1.23$, $\beta = -.296$, $t(2031) = -15.98$, $p < .001$, and MFT scenarios were less severe than naturalistic scenarios, $b = -.801$, $\beta = -.192$, $t(2031) = -10.38$, $p < .001$. Results also revealed a significant Content × Source interaction, $b = -.910$, $\beta = -.109$, $t(2028) = -6.02$, $p < .001$. Simple slope analysis (Aiken & West, 1991) showed that, consistent with our prediction, the effect of content on severity was much greater for MFT scenarios, $b = -1.69$, $\beta = -.405$, $t(962) = -15.17$, $p < .001$, than for naturalistic scenarios, $b = -.778$, $\beta = -.187$, $t(962) = -7.47$, $p < .001$ (controlling for politics). Interpreted differently, the relationship between scenario source and severity was stronger for impurity, $b = -1.26$, $\beta = -.301$, $t(962) = -11.23$, $p < .001$, than for harm, $b = -.346$, $\beta = -.083$, $t(962) = -3.65$, $p < .001$. In short, while the MFT scenarios were somewhat less severe overall, this was particularly true of the MFT impurity scenarios.

Weirdness

Overall, impurity scenarios were weirder than harm scenarios, $b = .850$, $\beta = .193$, $t(2031) = 10.71$, $p < .001$, and MFT scenarios were weirder than naturalistic scenarios, $b = 1.66$, $\beta = .377$, $t(2031) = 20.92$, $p < .001$. The Content × Source interaction was also significant, $b = 1.16$, $\beta = .132$, $t(2028) = 7.43$, $p < .001$. Simple slope analysis revealed that, as predicted, the effect of content on weirdness was considerably greater for MFT scenarios, $b = 1.43$, $\beta = .325$, $t(962) = 12.29$, $p < .001$, than for naturalistic scenarios, $b = .269$, $\beta = .061$, $t(962) = 2.86$, $p = .004$. Correspondingly, the relationship between scenario source and severity was greater for impurity scenarios, $b = 2.24$, $\beta = .509$, $t(962) = 20.32$, $p < .001$, than for harm scenarios, $b = 1.08$, $\beta = .245$, $t(962) = 9.72$, $p < .001$. In short, while the MFT scenarios were somewhat weirder overall, this was particularly true of the MFT impurity scenarios.

Harm

Overall, harm scenarios were more harmful than purity scenarios, $b = -1.89$, $\beta = -.428$, $t(2031) = -24.17$, $p < .001$, and MFT scenarios were less harmful than naturalistic scenarios,

$b = -.765$, $\beta = -.173$, $t(2031) = -9.78$, $p < .001$. Results also revealed a significant Content × Source interaction, $b = -.951$, $\beta = -.108$, $t(2028) = -6.19$, $p < .001$. Simple slope analysis showed that the effect of content on harmfulness was greater for MFT scenarios, $b = -2.37$, $\beta = -.535$, $t(962) = -21.68$, $p < .001$, than for naturalistic scenarios, $b = -1.41$, $\beta = -.320$, $t(962) = -12.88$, $p < .001$. Expressed differently, the association between scenario source and harmfulness was stronger for impurity scenarios, $b = -1.24$, $\beta = -.281$, $t(962) = -10.77$, $p < .001$, than for harm scenarios, $b = -.290$, $\beta = -.067$, $t(962) = -3.10$, $p = .002$.

Impurity

Overall, harm scenarios were *more impure* than impurity scenarios, $b = -.600$, $\beta = -.148$, $t(2031) = -7.76$, $p < .001$, and MFT scenarios were less impure than naturalistic scenarios, $b = -.710$, $\beta = -.176$, $t(2031) = -9.19$, $p < .001$. The Content × Source interaction was also significant, $b = -.759$, $\beta = -.094$, $t(2028) = -5.01$, $p < .001$. Simple slope analysis showed that effect of content on impurity was much greater for MFT scenarios, $b = -.979$, $\beta = -.242$, $t(962) = -8.51$, $p < .001$, than for naturalistic scenarios, $b = -.221$, $\beta = -.055$, $t(962) = -2.20$, $p = .03$. Analyzed differently, the relationship between scenario source and impurity was stronger for impurity scenarios, $b = -1.09$, $\beta = -.270$, $t(962) = -9.86$, $p < .001$, than for harm scenarios, $b = -.331$, $\beta = -.082$, $t(962) = -3.51$, $p < .001$. In other words, as in Study 1, scenarios describing harm were more likely to elicit perceptions of impurity than scenarios engineered to capture impurity. This effect was greatest for MFT impurity scenarios, suggesting that the naturalistic impurity scenarios assess (MFT defined) impurity better than MFT impurity scenarios.

Discussion

Demonstrating sampling bias, MFT impurity scenarios were more weird and less severe than naturalistic impurity scenarios. Naturalistic purity violations were still somewhat less severe and weirder than harm violations—as might be predicted by modular morality—but such differences are also consistent with more parsimonious domain-general morality: A harm-based prototype (Gray & Schein, 2012) suggests that harm violations better represent the category "immoral" (i.e., are more severe) and are more typical (i.e., are less weird). Most importantly, these results suggest that any potential differences in weirdness and severity are exaggerated by popular MFT scenarios. As in Study 1, ratings of harm and impurity were highly correlated, $r(20) = .87$, $p < .001$, again casting doubt on their distinctness.

Study 3: Severity and Weirdness, Act and Character

Studies 1 and 2 revealed that commonly used moral scenarios confound moral content with severity and weirdness. Perhaps the key question is why should we worry about severity and weirdness sampling bias in scenarios? The answer is that these dimensions—and not moral content per se—could be driving different patterns of moral judgment, giving the illusion of specialized mechanisms while actually supporting the importance of domain-general dimensions.

For example, Uhlmann and Zhu (Study 2a, 2014) found that harm violations (e.g., theft) were rated as more immoral but less indicative of poor moral character than purity violations (e.g., dog-eating and chicken sex). Here, we test whether weirdness and severity account for the apparent pattern of harm = acts and purity = character. We selected naturalistic harm and impurity scenarios closely matched on severity and weirdness—adultery and assault (See Figure 3)—and independently manipulated severity and weirdness in a factorial design.

Given that severity measures overall moral magnitude, we expected that it would predict judgments of both acts and character. We also hypothesized that weirdness would uniquely influence judgments of character because of the link between counternormativity and dispositional attributions (Jones & Davis, 1965; Pizarro & Tannenbaum, 2011).

Method

Four hundred and seventy-eight participants were recruited through mTurk. Thirty-three failed to finish and 34 failed attention checks, leaving 411 (55% male, $M_{\text{age}} = 31$, 60% liberal).

Participants were randomly assigned scenarios in a 2 (content: harm vs. impurity) × 2 (severity: severe vs. mild) × 2 (weirdness: weird vs. typical) × 2 (evaluation type: act vs. character) mixed-factorial design, in which content, severity, and weirdness were between-subjects variables and evaluation type was a within-subjects variable. In the harm scenarios, participants were asked to "Imagine a man [slaps someone on the face (severe)/steps on someone's foot (mild)]." In the impurity scenarios, participants were asked to "Imagine a man [French kisses and gropes (severe)/dances with (mild)] someone who is not his wife." For the weird conditions, the sentence concluded, "after painting himself red and putting on a cape made of old human hair." Following Uhlmann and Zhu (2014), participants provided act evaluations ("Is this behavior morally wrong?") or character evaluations ("Does this person have poor moral character?") using a 7-point scale from (1) *Definitely not* to (7) *Definitely yes*. Severity, weirdness, and political affiliation were rated as in Study 2.

Results

Manipulation checks confirmed that severe scenarios ($M = 3.82$, $SD = 1.64$) were more severe than mild scenarios ($M = 2.07$, $SD = 1.41$), $t(409) = 11.55$, $p < .001$ and that weird scenarios ($M = 6.30$, $SD = 1.32$) were weirder than typical scenarios ($M = 3.08$, $SD = 1.83$), $t(409) = 19.79$, $p < .001$. Overall, scenarios were well matched in terms of severity (harm $M = 3.04$, $SD = 1.61$; impurity $M = 2.96$, $SD = 1.92$), $t(409) < 1$, and weirdness (harm $M = 4.60$, $SD = 2.26$; impurity $M = 4.32$, $SD = 2.29$), $t(409) = 1.25$, $p = .21$). See Supplementary Table S9 for the means for each condition.

To test for the role of severity and weirdness, a 2 (content) × 2 (severity) × 2 (weirdness) × 2 (evaluation type) mixed-model analysis of variance (ANOVA) was estimated. The ANOVA revealed no main effect of moral content, $F(1, 402) = .679$, $p = .41$, or evaluation type, $F(1, 402) = .684$, $p = .41$. Significant main effects of severity, $F(1, 402) = 148.74$, $p < .001$, and weirdness, $F(1, 402) = 36.31$, $p < .001$, indicated that severe and weird violations were judged more harshly across content and evaluation types.

The main effect of weirdness was qualified by a significant Weirdness × Evaluation type interaction, $F(1, 402) = 5.81$, $p = .02$. Consistent with our predictions, the effect of weirdness was larger for character evaluations, $F(1, 196) = 19.52$, $p < .001$ ($M_{\text{weird}} = 4.68$, $M_{\text{typical}} = 3.64$), than for act evaluations, $F(1, 199) = 9.06$, $p = .003$ ($M_{\text{weird}} = 4.57$, $M_{\text{typical}} = 3.86$).

The Severity × Evaluation type interaction was only marginally significant, $F(1, 402) = 3.35$, $p = .07$, supporting our prediction that severity is an important determinant of both evaluation types. Simple effects analysis revealed no effect of evaluation type for mild violations, $F(1, 190) = 0.50$, $p = .48$ ($M_{\text{act}} = 3.18$, $M_{\text{character}} = 3.23$), and a marginal effect of evaluation type for severe scenarios, $F(1, 213) = 3.56$, $p = .06$ ($M_{\text{act}} = 4.97$, $M_{\text{character}} = 5.08$). This finding may suggest that severe violations are, due to their rarity, also quite diagnostic of moral character (Uhlmann, Pizarro, & Diermeier, 2015).

Analyses revealed a significant Severity × Weirdness interaction, $F(1, 402) = 10.94$, $p = .001$. The effect of weirdness was larger for mild violations, $F(1, 190) = 38.55$, $p < .001$ ($M_{\text{weird}} = 3.98$, $M_{\text{typical}} = 2.63$), than for severe violations, $F(1, 213) = 4.17$, $p = .04$ ($M_{\text{weird}} = 5.27$, $M_{\text{typical}} = 4.88$), consistent with a model in which severity is the most important feature of a moral violation, and weirdness is free to play an important role in moral judgment formation only when severity is relatively low.

Importantly, the Content × Evaluation type interaction was not significant, $F(1, 402) = 1.72$, $p = .19$. Thus, when scenarios are matched for severity and weirdness, we find no support for a link between purity per se and evaluation type, previously documented elsewhere as evidence of modular morality (e.g., Uhlmann & Zhu, 2014).

The three-way Content × Severity × Weirdness interaction was significant, $F(1, 402) = 10.39$, $p = .001$. The Severity × Weirdness interaction reported earlier was particularly pronounced for purity violations: The effect of weirdness was large and significant for mild violations, $F(1, 95) = 37.53$, $p < .001$ ($M_{\text{weird}} = 4.08$, $M_{\text{typical}} = 2.24$), but not for severe violations, $F(1, 106) = 0.03$, $p = .86$ ($M_{\text{weird}} = 5.07$, $M_{\text{typical}} = 5.13$). For harm violations, the effect of weirdness was more comparable for mild violations, $F(1, 95) = 7.56$, $p = .007$ ($M_{\text{weird}} = 3.89$, $M_{\text{typical}} = 3.02$), and severe violations, $F(1, 107) = 12.65$, $p = .001$ ($M_{\text{weird}} = 5.47$, $M_{\text{typical}} = 4.62$). As in Study 2, this suggests that severity and weirdness likely do not account for all differences between harm and impurity scenarios. However, such residual differences are not necessarily suggestive of modularity, as other domain-general dimensions may be relevant. For example, attributional ambiguity (i.e., multiple plausible motives; Snyder, Kleck, Strenta, & Mentzer, 1979) may be higher in the case of slapping (e.g., playing around, taking offense, and displaying dominance) than French kissing. The other three-way interactions, and the four-way interaction, were not statistically significant (all *F*s < 1; all *p*s > .5).

Next, the ANOVA was reestimated including political orientation (standardized) as a covariate. The main effect of political orientation was not significant, $F(1, 402) = .795$, $p = .37$, nor was the Politics × Evaluation type interaction, $F(1, 402) = 2.06$, $p = .15$. As the supplementary materials detail (Note 3), the effects reported earlier were virtually unchanged with the inclusion of politics.

Discussion

Using severity- and weirdness-matched harm and impurity scenarios, we found that manipulations of these dimensions—not moral content—predicted act and character judgments, arguing against the intrinsic importance of moral content and therefore modular accounts (Uhlmann & Zhu, 2014).

General Discussion

Across three studies, we found evidence for sampling bias in moral psychology. In popular MFT scenarios, moral content (harm vs. impurity) is confounded with severity and weirdness (Study 1). Although some severity and weirdness differences persist in naturalistic scenarios, MFT scenarios inflate these differences (Study 2). This sampling bias is practically important because domain-general severity and weirdness can explain effects previously ascribed to moral modules, such as differences between act and character judgments (Study 3).

Furthermore, Studies 1 and 2 document very high correlations between harm and impurity, *r*s > .86, suggesting that lay intuitions do not reflect the sharp content boundaries hypothesized by modularists (Haidt & Joseph, 2005). Considering the low reliability of moral judgments within MFT-defined content areas of harm ($\alpha = .51$) and impurity ($\alpha = .75$) revealed by past work (Graham et al., 2011, p. 372), the Spearman attenuation-corrected correlation between them ($r = 1.0$)[2] raises doubts about *any* empirical separation between harm and impurity. If anything, MFT impurity scenarios are rated as *less* impure than harm scenarios, posing a substantial problem for MFT.

Compared to MFT scenarios, naturalistic impurity scenarios are less biased and better reflect lay intuitions, but we must acknowledge that they are still somewhat weirder and less severe than harm scenarios. Do these differences reflect intrinsic characteristics of a distinct and specialized impurity mechanism? We believe that this is unlikely for two reasons. First, these data suggest that the construct of "impurity" is neither distinct from harm (Studies 1 and 2) nor specialized (Study 3) nor even best predicted by "impurity" scenarios (Studies 1 and 2). Second, a harm-as-prototype account (Gray & Schein, 2012) suggests that *any* immoral acts in which harm is less obvious than direct physical harm (e.g., assault, murder) should be seen as less severe (less "immoral") and less typical (less prototypical). Based upon naturalistic scenarios and the overlap between harm and impurity, perhaps we can simply define participants' understanding of impurity as "(perceived) harm involving sex." This does not require a distinct moral judgment mechanism any more than listening to "music involving saxophones" requires a distinct music-listening mechanism.

Of course, MFT scenarios have revealed some differences between "harm" and "impurity" scenarios, but these may reflect specific scenario details (e.g., bestiality, incest, bizarre surgery) rather than broader moral "foundations." Even political differences revealed by MFT scenarios may merely reflect these low-level details. Conservatives may seem more concerned about impurity when assessed through sex and religion (Graham et al., 2009), but liberals are likely more concerned with nutritional and environmental contamination (Feinberg & Willer, 2013). Indeed, research shows that political conservatives are more sensitive to impurity concerns when asked about anal sex but less sensitive when asked about eating fast food (Jarudi, 2009).

We acknowledge that only a small subset of moral scenarios were investigated, but scenarios and other self-report measures constitute the method "most widely used by far" in the literature (Graham et al., 2012, p. 47). Studies with these measures have been cited as evidence for modularity within studies of anger versus disgust (Hutcherson & Gross, 2011; Rozin, Lowery, Imada, & Haidt, 1999; Russell & Giner-Sorolla, 2011; Schnall, Haidt, Clore, & Jordan, 2008; Seidel & Prinz, 2013; but see Chapman & Anderson, 2014), self versus other (Chakroff, Dungan, & Young, 2013), murder versus suicide (Rottman, Kelemen, & Young, 2014; but see Gray, 2014), intention (Young & Saxe, 2011), psychopathy (Glenn, Iyer, Graham, Koleva, & Haidt, 2009), and differential brain activations (Parkinson et al., 2011). Given the newly revealed role of severity and weirdness in act and character judgments, replication of these other effects may also reveal support for domain-general—and more parsimonious—accounts of moral cognition (see also DeScioli, Asao, & Kurzban, 2012). Even recent attempts at "standardized" moral foundation scenarios fail to control for weirdness and severity (Clifford, Iyengar, Cabeza, & Sinnott-Armstrong, 2015).

Future studies must ensure that scenarios manipulating content are equated on severity and typicality, and ensure that impurity scenarios activate impurity concerns, but *not* harm concerns, and vice versa for harm scenarios. This may prove challenging, given the overlap between harm and impurity revealed here, but independent activation is an essential feature of separate mechanisms (Carey, 1995). More broadly, these results suggest that future studies should guard against sampling bias and confounds, enduring issues in both election prediction and moral psychology.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Participants also suggested many of the same violations (e.g., rape and adultery) for both harm and impurity.
2. Calculated as $r_{x'y'} = r_{xy} / \sqrt{(\alpha_{xx} \cdot \alpha_{yy})}$ using these previously published reliabilities in conjunction with correlations between ratings of harm and impurity Studies 1 and 2.

Supplemental Material

The online supplemental material is available at http://spps.sagepub.com/supplemental.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review*. Advanced online publication. doi:10.1177/1088868314566683

Carey, S. (1995). On the origin of causal understanding. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate*. (pp. 268–308). New York, NY: Oxford University Press.

Chakroff, A., Dungan, J., & Young, L. (2013). Harming ourselves and defiling others: What determines a moral domain? *PloS One*, *8*, e74434.

Chapman, H. A., & Anderson, A. K. (2014). Trait physical disgust is related to moral judgments outside of the purity domain. *Emotion*, *14*, 341–348. doi:10.1037/a0035120

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, 1–21. doi:10.3758/s13428-014-0551-2

DeScioli, P., Asao, K., & Kurzban, R. (2012). Omissions and byproducts across moral domains. *PLoS ONE*, *7*, e46963. doi:10.1371/journal.pone.0046963

Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, *24*, 56–62. Retrieved from http://doi.org/10.1177/0956797612449177

Glenn, A. L., Iyer, R., Graham, J., Koleva, S., & Haidt, J. (2009). Are all types of morality compromised in psychopathy? *Journal of Personality Disorders*, *23*, 384–398. doi:10.1521/pedi.2009.23.4.384

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2012). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. doi:10.1016/B978-0-12-407236-7.00002-4

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*, 1029–1046. doi:10.1037/a0015141

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

Gray, K. (2014). Harm concerns predict moral judgments of suicide: Comment on Rottman, Young and Keleman (2014). *Cognition*, *130*, 217–226.

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*, 1–19. doi:10.1007/s13164-012-0112-5

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*, 1600–1615.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*, 101–124. doi:10.1080/1047840x.2012.651387

Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion*, *7*, 853–868. doi:10.1037/1528-3542.7.4.853

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98–116. Retrieved from http://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*, 55–66.

Haidt, J., & Joseph, C. (2005). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. *3*, pp. 367–391). New York, NY: Oxford University Press.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The Innate Mind*, *3*, 367–392.

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*, 613–628.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social–functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, *100*, 719.

Impure [Def. 2]. (n.d.) *Oxford English Dictionary*. Retrieved June 3, 2015, from http://www.oed.com/view/Entry/92939?rskey=MrDRaF&result=1&isAdvanced=false#eid

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science*, *7*, 496–503.

Jarudi, I. N. (2009). *Everyday morality and the status quo: Conservative concerns about moral purity, moral evaluations of everyday objects, and moral objections to performance enhancement* (Doctoral Dissertation). Yale University, New Haven, CT.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. *2*, pp. 219–266). New York, NY: Academic Press.

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. Hoboken, NJ: John Wiley.

Murphy, G. L. (2004). *The big book of concepts*. Cambridge: MIT Press.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, *23*, 3162–3180. doi:10.1162/jocn_a_00017

Pizarro, D. A., Inbar, Y., & Helion, C. (2011). On disgust and moral judgment. *Emotion Review*, *3*, 267–268. doi:10.1177/1754073911402394

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington, DC: APA Press.

Pliner, P., & Pelchat, M. L. (1991). Neophobia in humans and the special status of foods of animal origin. *Appetite*, *16*, 205–218. doi:10.1016/0195-6663(91)90059-2

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57–75. doi:10.1037/a0021867

Rottman, J., Kelemen, D., & Young, L. (2014). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition*, *130*, 217–226. doi:10.1016/j.cognition.2013.11.007

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*, 574–586. doi:10.1037/0022-3514.76.4.574

Russell, P. S., & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion*, *11*, 233–240. doi:10.1037/a0022598

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*, 1096–1109.

Seidel, A., & Prinz, J. (2013). Sound morality: Irritating and icky noises amplify judgments in divergent moral domains. *Cognition*, *127*, 1–5.

Shweder, R. A., & Haidt, J. (1994). The future of moral psychology: Truth, intuition, and the pluralist way. In B. Puka (Ed.), *Reaching out: Caring, altruism, and prosocial behavior* (pp. 336–341). New York, NY: Garland. Retrieved from http://humdev.uchicago.edu/publications/shweder/ShwederFutureMoralPsychology.pdf

Snyder, M. L., Kleck, R. E., Strenta, A., & Mentzer, S. J. (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology*, *37*, 2297–2306. doi:10.1037/0022-3514.37.12.2297

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

Uhlmann, E. L., & Zhu, L. [Lei] (2014). Acts, persons, and intuitions person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science*, *5*, 279–285. doi:10.1177/1948550613497238

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*, 202–214. doi:10.1016/j.cognition.2011.04.005

Author Biographies

Kurt Gray is an assistant professor of psychology who studies mind perception and morality.

Jonathan E. Keeney is a PhD student who explores the interplay of moral cognition and real-world decision making.