Judgment and Decision Making, Vol. 11, No. 5, September 2016, pp. 496–508

Representations of moral violations: Category members and associated features

Justin F. Landy*

Abstract

I present a novel way to conceptualize Turiel and colleagues' Social Domain Theory (SDT), and Haidt and colleagues' Moral Foundations Theory (MFT), as theories of how concepts of moral violations are mentally represented. I argue that SDT is best viewed as a theory of the features that are associated with concepts of moral violations, including wrongness, generalizability across cultures, and intrinsic harmfulness, and that MFT, in contrast, is best viewed as a theory of individual differences in what kinds of acts are categorized as moral violations (i.e., of category membership). This perspective generates a novel prediction: the same individual difference variables that predict variation in moral values according to MFT should predict ascription of the features predicted by SDT. That is, judgments of wrongness, generalizability, and intrinsic harmfulness should covary with the same predictors as do endorsed moral values, specifically, political orientation and analytic thinking. Three studies supported this hypothesis.

Keywords: moral judgment, Social Domain Theory, Moral Foundations Theory, individual differences

1 Introduction

Is stealing another person's wallet immoral? If you are like most people, you probably think that it is. Assuming that you do, I might further ask you, what does it mean to say that theft is immoral? You might reply that you mean that theft is wrong, that people should not steal each other's wallets. You might also, after a bit of further probing, say that theft is always or nearly always wrong – one should not steal another person's wallet, even if one lives in a lawless land where wallet theft is the norm. If you are particularly sophisticated (or have read a lot of moral psychology research), you might further point out that this is because stealing a wallet always hurts the wallet's owner, and this is true regardless of whether a person or culture believes it to be.

The preceding questions can be thought of as questions about the categorization of concepts, and can be rephrased to reflect this. Is the concept WALLET THEFT a member the category MORAL VIOLATIONS?[1] Separately, what features do we associate with the concept WALLET THEFT, as a member this category? The former is a question of category membership: which concepts are members (or exemplars) of the category MORAL VIOLATIONS? The latter is a question of the common features[2] associated with concepts that are considered to belong in this category, whatever those concepts may be (see Berniūnas, Dranseika & Sousa, 2016, for a similar distinction between "conviction" and "content" in moral judgment, and Shweder & Much, 1991, for an earlier, though largely unexplored, distinction between "content" and "form").

It is important to note that the members of the category MORAL VIOLATIONS and the features associated with these category members are both aspects of people's mental representations of the category MORAL VIOLATIONS, as opposed to the cognitive process by which people categorize acts as belonging to this category. Representations of concepts and categories are distinct from the processes by which we manipulate those representations. For instance, the category BALLS contains certain category members: BASKETBALLS, GOLF BALLS, BASEBALLS, etc., and certain features are typically associated with members of this category, such as "round", "used in playing sports", "man-made", and so forth. Both the membership of the category, and the features associated with the concepts that belong to it, are aspects of how we represent BALLS, and are distinct from the processes by which we categorize novel concepts like BOCCE BALL as

I am grateful to Daniel Bartels, Edward Royzman, and Sydney Scott for their insightful comments on earlier versions of this manuscript.

*Center for Decision Research, University of Chicago Booth School of Business, 5807 S Woodlawn Avenue, Chicago, IL, 60637, USA. Email: justinlandy@chicagobooth.edu.

[1] Throughout this paper, I will follow convention by using SMALL CAPS to refer to specific concepts and categories. Thus, "MORAL VIOLATIONS" refers to the mental representation of a category, while "moral violations" refers to the various concepts that are considered to be exemplars of that category.

[2] The phrase "common features" is not meant to imply features that are necessary and sufficient for an act to be categorized as a moral violation, as in the classical definition of concepts. The argument that I will advance here is compatible with the classical definition, but does not require it. When I argue that the acts that are categorized as moral violations are ascribed certain features, I mean that they generally tend to be thought to possess those features, more so than other kinds of acts that are not categorized as members of MORAL VIOLATIONS, consistent with prototype and exemplar theories of concepts (see Medin & Rips, 2005, for a review of competing theories of how concepts should be conceptualized).

belonging to the category. Similarly, the membership of the category MORAL VIOLATIONS is a property of how that category is represented in memory, and certain features are associated with members of this category, but both of these aspects of representation are separate from the process by which we categorize acts as moral violations.

Moral psychologists have written extensively on the topic of process, and, in particular, whether the categorization of an act as a moral violation is a reasoned, thoughtful process (e.g., Royzman, Landy & Goodwin, 2014), an automatic, intuitive process (e.g., Haidt, 2001), or can involve both intuitive and deliberate processing in important ways (e.g., Greene, Sommerville, Nystrom, Darley & Cohen, 2001). This is a separate issue from the questions at hand here, which concern representations. The process by which acts are categorized as moral violations could be reasoned, intuitive, or some combination of both – any of these sorts of processes could plausibly produce the kinds of cognitive representations of moral violations for which I will argue.

The distinction between the membership of a category and the features ascribed to its members suggests that the features associated with acts categorized as moral violations could exhibit widespread agreement, while the specific acts that are categorized as moral violations could exhibit predictable individual differences. I will present evidence that this is indeed the case, and, in doing so, attempt to reconcile two influential theories of moral judgment that have typically been viewed as mutually opposed to one another: Social Domain Theory (SDT; Nucci & Nucci, 1982; Nucci & Turiel, 1978; Smetana, Jambon & Ball, 2014; Turiel, 1983, 2002, 2014) and Moral Foundations Theory (MFT; Graham, Haidt & Nosek, 2009; Haidt, 2012; Haidt & Graham, 2007; Haidt & Joseph, 2004). In short, I will argue that SDT is best thought of as a theory of features associated with concepts, whereas MFT is best thought of as a theory of category membership, and that this way of thinking not only allows for reconciling the two theories, but also generates novel, testable predictions.

1.1 Two theories of moral judgment

Social Domain Theory originated as a theory of moral development, and went on to become one of the most influential theories of moral judgment. A central claim of SDT is that counter-normative acts can be deemed "wrong" in two distinct ways, by violating moral rules or social conventions (Turiel, 1983, 2002). In the language of categorization, proscribed acts can be placed into two distinct categories, MORAL VIOLATIONS and CONVENTIONAL VIOLATIONS, the members of which have different features associated with them. Specifically, moral violations, as compared to conventional violations, are considered to be more wrong, and are seen as being enforceable regardless of culture or consensus – that is, they are considered to be *generalizable* to all times and places, and are thus seen as being less dependent on authority or socially endorsed rules for their normative force. These differences between morality and convention have been extensively studied using the classic paradigm known as the moral-conventional distinction task, in which participants indicate the wrongness of an act, and whether it would be wrong under a normative system where it is permitted. Paradigmatically moral violations are more likely to be judged wrong than conventional violations, even in normative contexts where they are not against the rules.

There is another, less frequently discussed distinction between moral and conventional violations in SDT, as well: moral violations intrinsically cause harm[3] to others (see, e.g., Turiel, 1983, p. 35, p. 221; see also Davidson, Turiel & Black, 1983; Haidt, 2008; Royzman, Landy et al., 2014). Indeed, this is why their wrongness cannot be nullified by authority or consensus – punching someone in the face without provocation will *always* cause harm, regardless of the cultural milieu in which the assault takes place. Conventional violations, on the other hand, may also cause harm, but only as a function of the culture in which they occur. In the United States, for instance, eating a steak with one's hands in an upscale restaurant would almost certainly be deemed disruptive and offensive by other patrons, but one can imagine cultures where this is the norm and offends no one (Huebner, Lee & Hauser, 2010). In other words, the harm is not intrinsic to the act itself, but results from the act occurring within a particular normative context.

Note that all of the above claims concern the features associated with concepts: those concepts that belong to the category MORAL VIOLATIONS, as compared to those that belong to CONVENTIONAL VIOLATIONS, are considered to be more wrong, more generalizable, and more intrinsically harmful. SDT also makes claims about what concepts are considered to be members of the category MORAL VIOLATIONS, but recent research has found these claims wanting. Specifically, the moral domain is said to concern "justice, rights, and welfare" (Turiel, 1983, p. 3), with all other types of violations (e.g., counter-normative sexual acts, addressing authority figures informally, etc.) falling outside of the moral domain.

So, according to SDT, only violations pertaining to justice, rights, and welfare ought to be considered especially wrong, generalizable, and intrinsically harmful, but recent research suggests that this is not the case for all people. Indeed, there is good evidence that some people consider certain counter-normative sexual acts, at least, to be generalizable to other normative contexts, in the way that paradigmatic moral violations are considered to be (Haidt & Hersh, 2001; Haidt, Koller & Dias, 1993; Royzman, Landy et al.,

[3] I use "harm" in a broad sense, essentially to mean "negative outcomes", rather than direct physical or emotional harm. This differs from the more narrow definition of harm meant by MFT's "care/harm" foundation, which I will refer to as "care" to avoid confusion.

2014).[4] Thus, the claims made in SDT regarding which types of acts are categorized as moral violations appear not to apply to all people, but little is known about how broadly applicable the claims about features associated with moral violations are. I propose that the claims about features of moral violations in SDT are generally correct – people do consider acts that make up the category MORAL VIOLATIONS to be more wrong, more generalizable, and more intrinsically harmful than acts outside of this category, but people vary in which acts they consider to be members of this category.

Moral Foundations Theory grew out of research showing that SDT's claims about category membership do not apply to all people. It is a much newer theory than SDT (introduced by Haidt and Joseph in 2004), but it has already spurred an enormous amount of research. The central claim of MFT is that there are (at least) five "foundations", or classes of virtues, that people moralize (Graham et al., 2009; Haidt, 2012; Haidt & Graham, 2007; Haidt & Joseph, 2004). The foundations of care/harm and fairness/cheating are concerned with preventing direct physical and emotional harm and promoting welfare, and with justice, rights, and fair outcomes, respectively. These are referred to as "individualizing foundations" because they promote individual autonomy and well-being, and they closely resemble the conception of the moral domain according to SDT, as being about "justice, rights, and welfare". The foundations of authority/subversion, loyalty/betrayal, and sanctity/degradation are concerned with respect for and obedience to legitimate authority, loyalty to important groups like one's family and nation, and bodily and spiritual purity, respectively. These are referred to as "binding foundations" because they are said to bind individuals into moral communities.

Since its introduction, MFT has generated a great deal of research on individual differences in espoused moral values. In particular, differences in political beliefs and analytic thinking both appear to predict what concerns people consider relevant to morality. Political conservatives are more likely to endorse statements of the binding foundations (e.g., "Respect for authority is something all children need to learn" [Authority], "I would call some acts wrong on the grounds that they are unnatural" [Sanctity]) than are political liberals (Graham et al., 2009, though see Davis et al., 2016, for a recent caveat). Endorsement of the binding foundations is also negatively related to individual differences in analytic thinking; both cognitive ability (i.e., intelligence) and a reflective cognitive style, measured by the Cognitive Reflection Test (CRT; Frederick, 2005) are negatively associated with explicit endorsement of the binding foundations as morally relevant (Pennycook, Cheyne, Barr, Koehler & Fugelsang, 2014), as is a dispositional preference for rational thinking (Garvey & Ford, 2014). MFT is primarily a theory of category membership. Violations of each foundation have their own associated features (e.g., violations of the sanctity foundation are often disgusting), but MFT has little to say about what features we *generally* ascribe to acts categorized as moral violations. Indeed, this claim that MFT is primarily about category membership seems quite consistent with a description of the project of MFT as being about "mapping the moral domain" (Graham et al., 2011). I propose, therefore, that while MFT captures important individual differences in the membership of the category MORAL VIOLATIONS, it can be informed by the claims about features associated with moral violations from SDT, as it has little to say on this topic, on its own.

[4] There is also some evidence that violations relating to justice, rights, and welfare are not always generalized (Kelly, Stich, Haley, Eng & Fessler, 2007). This issue remains unsettled (Sousa, 2009; Sousa, Holbrook & Piazza, 2009; Stich, Fessler & Kelly, 2009), so I will focus here on acts, such as sexual violations, that SDT would predict to be outside of the moral domain.

1.2 The present research

If SDT describes well the features that are associated with concepts of moral violations, and MFT describes well individual differences in what kinds of actions are considered to be members of the category MORAL VIOLATIONS, this leads to a novel prediction: the same individual differences that predict which foundations people endorse as morally important should predict ascription of the features predicted by SDT to violations of those foundations. That is, if the features associated with concepts of moral violations exhibit widespread regularity, while the membership of the category MORAL VIOLATIONS varies predictably, then judgments of an act's wrongness, generalizability, and intrinsic harmfulness should be predicted by the same individual difference variables that predict explicit endorsement of different moral foundations, such as political beliefs and analytic thinking. For instance, political conservatives, who more strongly endorse respect for authority as a moral virtue, would be expected to consider the "harmless" act of privately calling one's boss an "idiot" to be wrong, wrong regardless of culture or consensus (i.e., generalizable), and intrinsically harmful to others, more so than political liberals. If this were found to be the case, it would constitute evidence that SDT is a reasonable theory of the features that people ascribe to whatever acts they consider to be moral violations, and MFT correctly posits that the acts that are ascribed these features are not universally about "justice, rights, and welfare."

Other patterns of results would constitute evidence against this perspective. Suppose, for instance, that individual differences in political beliefs or analytic thinking were found to relate only to abstract endorsement of moral foundations, but not to judgments of contextualized violations (e.g., perhaps liberals condemn harmless cannibalism just as much as conservatives, but are just uncomfortable endorsing "unnaturalness" as relevant to moral judgments, in

the abstract). This would suggest that MFT is only a useful theory of espoused moral values, but not of membership of the category MORAL VIOLATIONS. Similarly, if conservatism or analytic thinking were found to relate consistently only to some kinds of moral judgments (e.g., wrongness), but not others (e.g., generalizability or intrinsic harmfulness), this would suggest that SDT does not correctly predict the features that people ascribe to members of the category MORAL VIOLATIONS, because the predicted features do not covary with the same predictors as one another, and endorsed values.

The only published study of which I am aware that examines comprehensively the relationship between political orientation and wrongness judgments of violations of the moral foundations was conducted by Clifford, Iyengar, Cabeza and Sinnott-Armstrong (2015), as part of the process of validating their Moral Foundations Vignettes, a database of vignettes describing behaviors that exemplify violations of the different moral foundations. As expected, ratings of the moral wrongness of the behaviors were related to political beliefs: conservatism predicted more severe wrongness ratings of loyalty, authority, and sanctity violations, but was mostly unrelated to ratings of care and fairness violations. This provides some evidence that the features of moral violations in SDT are predicted by the individual difference variables that predict foundation endorsement in MFT. However, greater wrongness is only one feature of moral violations in SDT. As of now, there is no evidence that ascriptions of other predicted features such as generalizability or intrinsic harmfulness are related to political beliefs. Moreover, while analytic thinking has been shown to predict wrongness and generalizability ratings of a handful of sexual offenses (Pennycook et al., 2014; Royzman, Landy et al., 2014), we do not know whether it predicts the wrongness, generalizability, and intrinsic harmfulness ascribed to violations of the binding foundations more broadly.

The present research examines in a more comprehensive fashion whether political beliefs and analytic thinking predict ascription of wrongness, generalizability, and intrinsic harmfulness to violations of the binding moral foundations. Based on my conception of SDT as a theory of associated features and MFT as a theory of category membership, I predicted that political conservatism and analytic thinking would predict wrongness and generalizability judgments of violations of the binding foundations (Studies 1 and 3) and ascriptions of intrinsic harmfulness to these violations (Studies 2 and 3).

2 Study 1

2.1 Method

Participants. I recruited approximately $N = 250$ participants for all three studies reported in this research, because correlations tend to stabilize as sample sizes approach 250 (Schönbrodt & Perugini, 2013), and prior research found that political conservatism correlated with generalizability judgments of sibling incest (a sanctity violation) at $r = .184$ (Royzman, Landy et al., 2014). A sample size of $N = 250$ should be able to detect an effect of this size at $\alpha = .05$ with statistical power of .90. This approximate sample size was determined before any data were collected.

Two hundred sixty-two participants located within the United States began this study on Amazon Mechanical Turk. Eleven did not complete the study and were therefore removed from the data, resulting in a final sample of $N = 251$.

Materials. The behaviors judged in this study were taken from the short form of the Moral Violations Database – Severity Equated (MVD-SE; Landy & Bartels, 2016), a set of brief descriptions of behaviors that pre-testing has shown are uniquely good exemplars of violations of each moral foundation. Importantly, the violations of each foundation are closely equated in their overall perceived wrongness (mean ratings: 4.99–5.04 on a 1–9 scale in validation studies with participants recruited through Mechanical Turk). I also included two completely non-moral actions from the MVD-SE to act as attention checks, for a total of 27 described behaviors (see the Supplement for complete stimuli). Participants also completed the thirty-item Moral Foundations Questionnaire (MFQ; Graham et al., 2011; available at moralfoundations.org). The MFQ measures explicit endorsement of the care, fairness, authority, loyalty, and sanctity foundations as morally important virtues (e.g., to what extent is "whether or not someone showed a lack of respect for authority" relevant when you decide whether something is right or wrong?), but does not include judgments of the sort of features of moral violations that the present research is concerned with. Participants also completed the three-item CRT (Frederick, 2005), and three syllogisms that require participants to overcome their prior beliefs when solving a logic problem (e.g., Markovits & Nantel, 1989; see Baron, Scott, Fincher, & Metz, 2014, for evidence that such items cohere with the CRT items as measures of careful, reflective thinking). The six-item measure of analytic thinking is presented in the Supplement.

Procedure. After giving informed consent, participants were presented with the 27 behaviors, and the MFQ, in counterbalanced order. The "moral relevance" section of the MFQ always preceded the "moral judgments" section (which is more about endorsing principles than judging actions, see Gray & Keeney, 2015), and the order of question presentation within each section was randomized for each participant. The order in which the 27 behaviors were presented was randomized for each participant, and each was presented on a separate page. Below each behavior, participants answered two questions, constituting a slightly modi-

Table 1: Descriptive statistics from Studies 1–3. Scale ranges are presented in brackets, and standard deviations are presented in parentheses.

Study 1	Political Conservatism [1–7]: 3.27 (1.67)				
	Analytic Thinking [0–6]: 3.79 (2.16)				
	Care	Fairness	Authority	Loyalty	Sanctity
MFQ [0–5]	3.55 (0.82)	3.58 (0.70)	2.51 (0.81)	2.32 (0.86)	2.01 (1.21)
Wrongness [1–9]	4.68 (1.38)	4.86 (1.44)	5.08 (1.55)	5.06 (1.29)	5.35 (2.00)
Generalizability [0–5]	2.34 (1.47)	2.47 (1.78)	2.56 (1.85)	2.34 (1.54)	2.32 (1.79)
Study 2	Political Conservatism [1–7]: 3.52 (1.73)				
	Analytic Thinking [0–6]: 3.41 (1.97)				
	Care	Fairness	Authority	Loyalty	Sanctity
Intrinsic Harm [1–9]	5.42 (1.49)	4.22 (1.47)	4.59 (1.62)	4.77 (1.38)	4.34 (1.77)
Study 3	Political Conservatism [1–7]: 3.42 (1.65)				
	Analytic Thinking [0–6]: 3.61 (2.00)				
	Care	Fairness	Authority	Loyalty	Sanctity
Wrongness [1–9]	5.25 (1.54)	5.19 (1.53)	5.42 (1.66)	5.54 (1.54)	5.68 (2.03)
Generalizability [1–9]	4.32 (1.74)	4.12 (1.87)	4.26 (1.88)	4.44 (1.88)	4.67 (2.14)
Intrinsic Harm [1–9]	5.77 (1.59)	5.04 (1.62)	5.13 (1.57)	4.80 (1.66)	4.93 (1.97)

fied form of the classic moral-conventional distinction task: "How wrong is this action?", answered on a 1-9 scale, and the generalizability probe. Following Royzman, Landy et al. (2014), this question described a hypothetical foreign country, Country A, where some time ago, the populace had all come together and decided that the described behavior was okay. Participants indicated whether the behavior would be wrong or not wrong, assuming that the person who did it was raised and lived in Country A. This question has been shown to correlate with other, alternative measures of generalizability (Royzman, Leeman & Baron, 2009).

After completing the moral judgments and MFQ, participants responded to the CRT and syllogisms. Lastly, participants responded to a brief demographics questionnaire, which included a single-item measure of political orientation (1 = strongly liberal, 7 = strongly conservative; this is the same scale used by Graham et al., 2009 in their original studies). Participants were then debriefed, thanked, and paid. No unreported measures were collected in any study reported in this paper.

2.2 Results

Preliminary analyses. Twenty-five participants failed at least one attention check, and were not included in the analyses below. The results remain substantively unchanged when these participants are included.

Internal reliabilities for wrongness judgments of each foundation ranged from acceptable to good (αs .60–.83), as did the internal reliabilities for generalizability judgments (αs .61–.80). I therefore averaged the continuous wrongness judgments, and summed the dichotomous generalizability judgments, for each foundation. Moreover, the three-item CRT and the three belief-bias syllogisms formed a reliable scale, α = .85, so I summed the number of correct responses to these six items to form a composite measure of analytic thinking. Descriptive statistics for key variables in Studies 1-3 are presented in Table 1. Estimated marginal means at ± 1 *SD* of conservatism and analytic thinking can be found in the Supplement.

Main analyses. There is some evidence that liberals tend to be more dispositionally analytic in their thinking (Deppe et al., 2015; Talhelm et al., 2015; though see Kahan, 2013, for countervailing results); however, in this study, conservatism and analytic thinking were not related, r = –.02, p = .73. Correlations between moral judgments, and political beliefs and analytic thinking, are presented in Table 2. Consistent with prior research, political conservatism was positively associated, and analytic thinking was negatively associated, with endorsement of the binding foundations as

Table 2: Correlations between individual difference variables and moral judgments in Study 1 (*df* = 224).

Political Conservatism	Care	Fairness	Authority	Loyalty	Sanctity
Wrongness	$-.08$	$.15^{*}$	$.27^{***}$	$.24^{***}$	$.23^{***}$
Generalizability	$-.08$	$.12\dagger$	$.15^{*}$	$.15^{*}$	$.24^{***}$
MFQ	$-.10$	$-.22^{**}$	$.34^{***}$	$.34^{***}$	$.37^{***}$
Analytic Thinking	Care	Fairness	Authority	Loyalty	Sanctity
Wrongness	$-.02$	.03	$-.14^{*}$	$-.18^{**}$	$-.32^{***}$
Generalizability	$-.10$	$-.05$	$-.13\dagger$	$-.22^{**}$	$-.33^{***}$
MFQ	$-.06$	.01	$-.14^{*}$	$-.17^{*}$	$-.23^{**}$

Note. $\dagger$ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$; MFQ = Moral Foundations Questionnaire.

morally relevant on the MFQ. More importantly, and consistent with the theory advanced here, conservatism was positively associated with, and analytic thinking was negatively associated with, wrongness ratings and generalizability judgments of violations of the binding foundations.[5] In other words, the same individual difference variables that predict endorsement of the binding foundations as morally relevant predict ascribing the features of moral violations predicted by SDT to violations of these foundations. These results remain essentially unchanged when statistically accounting for basic demographic variables (age, sex, race, education, and income), and when conservatism and analytic thinking are both included in the same linear model (see the Supplement for full regression tables). Similarly, endorsement of the individualizing foundations on the MFQ predicted wrongness and generalizability ratings of violations of these foundations, and endorsement of the binding foundations predicted wrongness and generalizability ratings of violations of these foundations (see the Supplement).

2.3 Discussion

As expected, political beliefs and analytic thinking both predicted wrongness and generalizability judgments of violations of the binding moral foundations. That is, the actions that people consider to belong to the category MORAL VIOLATIONS vary in ways that are consistent with MFT, but the features that people associate with the members of this category seem to be fairly consistent across individuals, and in line with the predictions of SDT.

[5]The only possible exception was the correlation between analytic thinking and generalizability judgments of authority violations, which was almost significant, $p = .061$ two-tailed.

3 Study 2

As mentioned above, one often overlooked claim of SDT is that acts that are morally (as opposed to conventionally) wrong are intrinsically harmful. This hypothesized property of moral violations is rarely, if ever, examined directly. Rather, it is typically argued that this is why prototypical immoral acts (i.e., violations of "justice, rights, and welfare" [Turiel, 2983, p. 3]) are generalized — the negative consequences of such acts cannot be removed by decree or consensus (Smetana et al., 2012). To my knowledge, no existing study has investigated individual differences in ascribing intrinsic harmfulness to acts, regardless of whether they "objectively" exhibit this property or not. If ascription of intrinsic harmfulness covaries with the same individual differences as wrongness and generalizability ascriptions and endorsement of different moral foundations, this would constitute further evidence that the features associated with concepts of moral violations are essentially consistent across individuals and consistent with SDT, even while the specific acts that are categorized as moral violations vary in important, predictable ways that are consistent with MFT.

3.1 Method

Participants. Two hundred seventy-five participants located within the United States began the study on Amazon Mechanical Turk. Two failed a mandatory "Captcha" verification, suggesting that they were "bot" programs, and 15 did not complete the study, leaving a final sample of N = 258. Participants from Study 1 could not take part in this study.

Materials and procedure. After consenting to participate, participants read instructions that explained the task, and the question that they would be answering. The instructions explained that "there are many kinds of behav-

iors that might be considered wrong, which differ from each other in lots of ways. One way in which they could differ is in whether they would negatively affect someone under any circumstances, or if any negative effects that they have could depend on the circumstances. For example, assaulting someone without provocation would always negatively affect that person. It is hard to imagine any circumstances where this would not be true. On the other hand, eating with your hands in a fancy restaurant might be considered offensive, and would be unpleasant for those around you, but it is not hard to imagine that some cultures would consider this to be perfectly acceptable and inoffensive." They went on to explain that "negative effects" could be "physical pain, emotional distress, economic loss, etc.", and that "what we want to know is which acts always have some negative effect, and which could have negative effects depending on the circumstances."

Participants were then presented with 27 descriptions of actions, each on a separate page. The first two were warm-up trials meant to acclimate participants to the task – a prototypical moral violation (making cruel remarks about a person's weight) and a prototypical conventional violation (attending a birthday party without bringing a gift). The remaining 25 were the violations from the MVD-SE used in Study 1, presented in a new randomized order for each participant. Each of the 27 actions was followed by a novel measure of intrinsic harm, "Is this action more like assaulting someone without provocation (it would always negatively affect someone, under any circumstances), or more like eating with your hands in a fancy restaurant (any negative effects depend on the circumstances)?". Responses were provided on a nine-point scale ranging from "More like assaulting someone without provocation" to "More like eating with your hands in a fancy restaurant" with the midpoint labeled "Somewhere in between". This novel measure was designed to capture the idea of intrinsic harm in straightforward, everyday language (Huebner, Dwyer & Hauser, 2009, proposed, but did not implement, a somewhat similar measure).

After responding to the 27 behavioral descriptions, participants completed the six-item measure of analytic thinking used in Study 1, followed by a demographics questionnaire that included a single-item measure of political orientation. After completing this, participants were debriefed, thanked, and paid.

3.2 Results

Preliminary analyses. Responses to the intrinsic harm items were reverse-scored such that higher numbers on the 1-9 scale indicate more perceived intrinsic harm. Internal reliabilities of intrinsic harm ratings for the five foundations ranged from borderline acceptable to good (αs .58–.76), so these ratings were averaged together to form five composite variables measuring perceived intrinsic harmfulness of violations of each foundation. As in Study 1, the six analytic thinking questions formed a reliable scale, α = .76, and were summed together to create a continuous measure.

Table 3: Correlations between individual difference variables and intrinsic harm ratings, in Study 2 (*df* = 256).

Foundation	Political Conservatism	Analytic Thinking
Care	$-.04$	.04
Fairness	.06	.02
Authority	$.18^{**}$	.06
Loyalty	$.14^{*}$	$-.15^{*}$
Sanctity	$.14^{*}$	$-.26^{***}$

Participants rated the prototypically moral warm-up item as significantly more intrinsically harmful (M = 6.67, SD = 1.98) than the prototypically conventional item (M = 2.34, SD = 1.53), $t(257) = 4.33$, $p < .001$, repeated-measures d = 1.78, indicating that they understood the task as intended.

Main analyses. As in Study 1, political conservatism and analytic thinking were not significantly correlated, $r(256) = -.07$, $p = .28$. Correlations between individual difference variables and intrinsic harm ratings for each of the moral foundations are presented in Table 3. As expected, political conservatism was positively related to ascribing intrinsic harmfulness to authority, loyalty, and sanctity violations, but not care and fairness violations. Similarly, analytic thinking was negatively related to ascribing intrinsic harmfulness to loyalty and sanctity violations, though, contrary to my predictions, not authority violations. The results are similar when statistically accounting for basic demographics (age, sex, race, education, and income), and when conservatism and analytic thinking are included in the same model (see the Supplement) for full regression tables).

3.3 Discussion

Perceived intrinsic harmfulness, a key feature of moral violations according to SDT, covaries with individual differences in political beliefs and analytic thinking in largely the same way as endorsement of moral concerns and wrongness and generalizability judgments, consistent with the theoretical perspective articulated here. The particular sorts of acts that people moralize vary predictably, while the features ascribed to those acts seem to be fairly consistent.

4 Study 3

The results of Studies 1 and 2 generally support my predictions. However, two criticisms might be leveled at the meth-

ods employed in these studies. First, judgments of wrongness, generalizability, and intrinsic harmfulness might all be measuring the same, single evaluation, i.e., that something is morally bad, or worthy of disapproval, rather than measuring the ascription of three distinct (though likely correlated) features. There is some reason to think that this is not the case: the mean correlation between wrongness and generalizability judgments across the five foundations in Study 1 was $r = .48$ (range: .38 - .75, all ps < .001); these measures are related to one another, but they do not seem to be redundant. However, we do not yet know how closely related intrinsic harmfulness judgments are to wrongness and generalizability judgments. To rectify this, all three measures were included in Study 3.

Second, the measure of intrinsic harmfulness used in Study 2 used concrete examples of intrinsically harmful ("more like assaulting someone without provocation") and non-intrinsically harmful actions ("more like eating with your hands in a fancy restaurant") as anchors on the response scale. This was intended to make the potentially unfamiliar notion of intrinsic harmfulness easy to understand, but it may have inadvertently led participants to evaluate the behaviors in that study on other dimensions on which these actions differ, such as moral wrongness. That is, the measure may have artificially conflated evaluations of wrongness with evaluations of intrinsic harmfulness. To remove this potential problem, the intrinsic harmfulness measure employed in Study 3 employed everyday language, but did not provide any concrete examples.

4.1 Method

Participants. Two hundred-seventy participants located within the United States began the study on Amazon Mechanical Turk. Eighteen did not complete the study, leaving a final sample of $N = 252$. Participants from Studies 1 and 2 could not take part in this study.

Materials and procedure. After consenting to participate, participants were told that they would read descriptions of various behaviors, then answer three questions about each one. They then read an explanation of the intrinsic harmfulness question: "The third question that you will answer is about the effects of each behavior. We want to know whether the behavior has bad effects (for someone other than the person who did it) 'built in', or if any bad effects would depend on something else." Similar to Study 2, they were then told that the bad effects of a behavior could be "physical pain, emotional distress, economic loss, or any other sort of harmful or damaging outcome." The instructions then continued, "For instance, some behaviors always hurt another person directly — these behaviors have bad effects built in. Other behaviors might be considered offensive or unpleasant by some people but not by others — these behaviors can have bad effects, but it depends on something other than the behavior itself. Some behaviors might fall somewhere in between." This turn of phrase, "having bad effects 'built in'" is intended to capture the idea of intrinsic harmfulness in straightforward, everyday language without employing explicit examples.

Participants were then presented with the same 27 behaviors as in Study 2 – two warm-up items, followed by the 25 critical behavioral descriptions – each followed by three questions. The wrongness question was identical to the one used in Study 1. The generalizability question was similar to that used in Study 1, except it employed a nine-point response scale identical to the wrongness question, for the sake of consistency across the three measures. The intrinsic harmfulness question read, "Does this behavior have bad effects built in, or would any negative outcomes depend on something other than behavior itself?" and employed a nine-point response scale ranging from "This behavior has bad effects built in" to "Any bad effects depend on something else", with the midpoint labeled "Somewhere in between". The order of presentation of the 25 critical behaviors was randomized for each participant.

After responding to the 27 behavioral descriptions, participants completed the six-item measure of analytic thinking used in Studies 1 and 2, followed by a demographics questionnaire that included a single-item measure of political orientation. After completing this, participants were debriefed, thanked, and paid.

4.2 Results

Preliminary analyses. Responses to the intrinsic harm items were reverse-scored such that higher numbers on the 1–9 scale indicate more perceived intrinsic harm. Internal reliabilities of wrongness, generalizability, and intrinsic harm ratings for the five foundations were acceptable (ranges .70–.82, .75–.85, and .61–.71, respectively), so these ratings were averaged together to form composite ratings for each foundation. As in the previous studies, the six analytic thinking questions formed a reliable scale, $\alpha = .78$, and were summed together to create a continuous measure.

Participants rated the prototypically moral warm-up item as significantly more wrong, generalizable, and intrinsically harmful than the prototypically conventional item, ts(252) > 12.79, ps < .001, repeated-measures ds > 1.64.

Main analyses. Wrongness and generalizability judgments were more highly correlated in Study 3 than in Study 1 (mean $r = .69$, range .59–.82, ps < .001), which may be an artifact resulting from the two questions using identical response scales. Wrongness and intrinsic harm (mean $r = .32$, range .19–.59, ps < .001) and generalizability and intrinsic harm (mean $r = .36$, range .21–.56, ps < .004) were

Table 4: Correlations between individual difference variables and moral judgments in Study 3 (*df* = 250).

Political Conservatism	Care	Fairness	Authority	Loyalty	Sanctity
Wrongness	−.03	.03	.17**	.16*	.30***
Generalizability	.05	.14*	.28***	.19**	.32***
Intrinsic Harm	−.04	−.07	.11†	.16*	.21**
Analytic Thinking	Care	Fairness	Authority	Loyalty	Sanctity
Wrongness	−.07	−.05	−.13*	−.21**	−.34***
Generalizability	−.16**	−.22***	−.30***	−.32***	−.38***
Intrinsic Harm	.12†	.03	−.13*	−.21**	−.24***

less strongly correlated. Overall, these measures are related, but seem not to be identical.

Unlike in Studies 1 and 2, conservatism and analytic thinking were negatively correlated, $r = -.30$, $p < .001$. This pattern of results across the three studies agrees with the existing literature, in that this association is sometimes observed, but may be somewhat weak and inconsistent (Deppe et al., 2015; Kahan, 2013; Talhelm et al., 2015).

Correlations between individual difference variables and judgments of wrongness, generalizability, and intrinsic harmfulness for each of the moral foundations are presented in Table 4. Replicating Studies 1 and 2, political conservatism was positively associated with all of these judgments for authority, loyalty, and sanctity violations,[6] but less related to judgments of care and fairness violations. Similarly, analytic thinking was negatively related to judgments of authority, loyalty, and sanctity violations, but less related to judgments of care and fairness violations. The results are similar when statistically accounting for basic demographics (age, sex, race, education, and income), though, because conservatism and analytic thinking were related in this sample, their independent predictive effects did not always persist when both predictors are entered in the same model (see the Supplement for full regression tables).

4.3 Discussion

Study 3 replicated the results of Studies 1 and 2 using an improved measure of intrinsic harmfulness, and demonstrated empirically that judgments of wrongness, generalizability, and intrinsic harmfulness are at least somewhat distinct, though all three covary with political beliefs and analytic thinking in the same ways as endorsement of binding moral foundations.

[6]The correlation between conservatism and intrinsic harm ratings of authority violations was marginally significant, $p = .076$.

5 General discussion

I have proposed that two putatively opposing theories of moral judgment, Social Domain Theory (SDT) and Moral Foundations Theory (MFT), actually have much to offer one another. Both of these theories can be viewed as concerning mental representations of moral violations, and I have argued that SDT is roughly correct as a theory of the features that are consistently associated with moral violations across individuals, while MFT captures the important, and predictable, individual differences in what actions are considered to be members of the category MORAL VIOLATIONS. This insight leads to the prediction that ascriptions of the features of moral violations predicted by SDT (specifically, wrongness, generalizability, and intrinsic harmfulness) should covary with the same individual difference measures that predict endorsement of different moral foundations, a prediction which has not been comprehensively tested previously. Three studies found that political conservatism and analytic thinking are related to judgments of wrongness, generalizability, and intrinsic harmfulness in the same way that they are to espoused moral principles. That is, conservatism was positively associated with these judgments of violations of authority, loyalty, and sanctity, while analytic thinking was negatively associated with them, supporting the contention that MFT captures important individual differences in the membership of the category MORAL VIOLATIONS, and SDT correctly predicts the features that people associate with the acts that belong to this category.

This same pattern of results is also observed at the level of individual stimuli (see the Supplement). Across the three studies, there are 15 authority/loyalty/sanctity stimuli x 6 total moral judgment items = 90 theoretically important item-level correlations between conservatism and moral judgment and 90 theoretically important item-level correlations between analytic thinking and moral judgment. Across so many statistical tests, *p*-values are largely uninformative, but the pattern of results is very consistent with the anal-

yses reported above: 88 out of 90 (97.7%) correlations between conservatism and moral judgments, and 84 out of 90 (93.3%) correlations between analytic thinking and moral judgments, were directionally consistent with my predictions. There was some variability in the strength of these correlations, as would be expected with a diverse set of stimuli, but the general pattern very much agrees with my hypotheses. It is also interesting to note that, in agreement with the analyses reported above, these item-level correlations tend to be strongest for sanctity violations. Of the stimuli used in this research, these are arguably the least "objectively" harmful, in that they contain nothing resembling a victim (or "moral patient", see Gray, Schein & Ward, 2014; Gray, Young & Waytz, 2012). It seems that the less "truly" harmful an action is, the less likely liberals and analytic thinkers are to perceive it as wrong, generalizable, and intrinsically harmful.

As noted above, while MFT has had little to say about the features that we generally ascribe to acts categorized as moral violations, it does seem to predict some features of various specific types of violations. Most notably, violations of sanctity are often thought to elicit disgust (Haidt, 2012; Horberg, Oveis, Keltner & Cohen, 2009; Rozin, Lowery, Imada & Haidt, 1999, though see Royzman, Atanasov, Landy, Parks & Gepty, 2014 for an alternative view). The theory advanced here says that ascription of the specific features predicted by SDT should covary with conservatism and analytic thinking. If I am correct that SDT is roughly accurate as a theory of features associated with concepts of moral violations, this implies that ascription of other features not predicted by SDT, such as disgustingness, should be less related to these individual difference variables. Some support for this comes from work by Royzman, Kim and Leeman (2015) who found no relationship between political beliefs and physical disgust reported in response to a vignette describing incest between siblings.

To provide a slightly more thorough (though by no means comprehensive) test, I recruited a new sample of 250 participants on MTurk and had them rate how "grossed out" (a lay term capturing the theoretical meaning of disgust, see Nabi, 2002) each of the five sanctity violations from Studies 1–3 made them feel. Ratings were almost significantly related to conservatism, $r(248) = .12$, $p = .054$, and were not reliably related to analytic thinking, $r(248) = –.10$. $p = .119$. Compare these relationships with the mean correlations between moral judgments of sanctity violations and these individual difference variables across Studies 1-3 (.24 and -.31, respectively) – the relationship with conservatism is twice as large for moral judgments, and the relationship with analytic thinking is over three times as large. The small correlation with conservatism agrees with prior research showing that conservatives are more easily disgusted than liberals (Inbar, Pizarro & Bloom, 2009), but it is clear that political beliefs more strongly relate to ascription of the features predicted by SDT than disgustingness, in agreement with the theoretical argument I have advanced.

To my knowledge, this article is the first empirical attempt to reconcile the claims of SDT and MFT, and the first research to explicitly couch them in the language of categorization and concepts. SDT is a theory of moral development, and MFT was developed by joining together insights from cultural anthropology and evolutionary psychology, so both theories have typically been divorced from research on categorization. This is likely why they have so often been seen as opposed, rather than as providing mutually informative insights on the nature of people's representations of moral violations.

5.1 Limitations and future directions

The clearest limitations of the present research concern the nature of the samples tested. Participants in all three studies were drawn from a single culture (the United States), so there is no evidence that the theory of moral concepts advanced here would generalize beyond this context. There is evidence that Westerners have a tendency to justify moral judgments with appeals to harm, even if harm appraisals play no role in producing the judgments, a tendency that may not be as prevalent in other cultures (Haidt, Björklund & Murphy, 2000; Haidt et al., 1993). However, more recent evidence suggests that reports of stubborn maintenance of moral condemnation despite belief that no harm has occurred (i.e., "moral dumbfounding") may have been overstated, and that people really do perceive harm in putatively "harmless" actions, consistent with my perspective here (Gray et al., 2014; Royzman et al., 2015). Of course, this tendency to perceive harm in all or nearly all actions categorized as moral violations may *itself* be restricted to Westerners or Americans, further complicating matters. It may be that intrinsic harmfulness (and perhaps generalizability as well) is not as consistently attributed to actions that violate valued moral principles in other cultural contexts as it is among Americans. With all of that said, the lion's share of the research on both SDT and MFT has been conducted with American samples, so the results presented here can still speak to both of these theories, as they have typically been studied. Cross-cultural research investigating the theoretical perspective put forth here is an important direction for future work.

A further concern relates to the use of the three original CRT items with samples recruited through Amazon Mechanical Turk, many of whom may have seen these items before. Indeed, the mean analytic thinking scores in my samples are fairly high, perhaps because of prior exposure. Note, however, that this effect would simply introduce noise into half of my analytic thinking measure, placing some

participants above their "true" analytic thinking score, and therefore making it less likely that any effect of this variable would be observed. That is, this potential issue with the measure biases against finding support for my hypotheses; the correlations reported here may underestimate the sizes of the key effects, but it seems unlikely that they overestimate them.

Lastly, there are, of course, other categories of behavior beyond the moral and the conventional. SDT also recognizes the "personal" domain, i.e., actions that are permissible, and prerogatives of the actor (e.g., Nucci, 1981; Tisak & Turiel, 1984). Moreover, Bicchieri has proposed a class of "social norms", as distinct from conventions (e.g., Bicchieri, 2005, 2010). Social norms are not necessarily internalized in the way that moral values are, and people will shirk social norms under conditions when there is little chance of detection, and when they do not expect others to adhere to them. An argument similar to the one advanced here may also apply to the distinctions between moral violations and personal prerogatives and between moral violations and social norm violations. For instance, liberals and analytic thinkers may consider a choice to violate the binding foundations to be more at the discretion of the actor (like personal actions) and/or more condemnable if other people are adhering to the norm (like violations of social norms). This would be quite consistent with the theoretical perspective articulated here, and investigating how consistent (or inconsistent) attributions of these features to different kinds of violations are across different people is an interesting and important direction for future research.

5.2 Conclusion

In this research, I have attempted to cast Social Domain Theory (SDT) and Moral Foundations Theory (MFT) in a new light, as theories of how people represent concepts of moral violations. By drawing a distinction between the features associated with these concepts, which appear to be fairly consistent across people, and the concepts that are categorized as belonging to the category MORAL VIOLATIONS, which vary in important and predictable ways, it is possible to reconcile these otherwise opposing theories. Advocates of SDT should concede that for at least some people, the moral domain goes well beyond "justice, rights, and welfare" (Turiel, 1983, p. 3), and that these differences in category membership are worthy of study, and proponents of MFT should acknowledge that the theory has had little to say about the features that are typically ascribed to actions considered to be moral violations. Examining these two theories in the context of the categorization of concepts allows them to mutually inform one another, and advances our understanding of the cognitive underpinnings of moral judgments.

References

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4,* 265–284.

Berniūnas, R, Dranseika, V., & Sousa, P. (2016). Are there different moral domains? Evidence from Mongolia. *Asian Journal of Social Psychology, 19,* 275–282.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge, UK: Cambridge University Press.

Bicchieri, C. (2010). Norms, preferences, and conditional behavior. *Politics, philosophy & economics*, *9*, 297–313.

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*, *47*, 1178–1198.

Davidson, P., Turiel, E., & Black, A. (1983). The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning. *British Journal of Developmental Psychology, 1,* 49–65.

Davis, D. E., Rice, K., van Tongeren, D. R., Hook, J. N., DeBlaere, C., Worthington, Jr., E. L., & Choe, E. (2016). The moral foundations hypothesis does not replicate well in Black samples. *Journal of Personality and Social Psychology, 110,* e23-e30.

Deppe, K. D., Gonzalez, F. J., Neiman, J. L., Jacobs, C., Pahlke, J., Smith, K. B., & Hibbing, J. R. (2015). Reflective liberals and intuitive conservatives: A look at the Cognitive Reflection Test and ideology. *Judgment and Decision Making, 10,* 314–331.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25–42.

Garvey, K. & Ford, T. G. (2014). Rationality, political orientation, and the individualizing and binding moral foundations. *Letters on Evolutionary Behavioral Science, 5,* 9–12.

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.

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.

Gray, K. & Keeney, J. E. (2015). Disconfirming Moral Foundations Theory on its own terms: Reply to Graham (2015). *Social Psychological and Personality Science, 6,* 874–877.

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.

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. (2008). Morality. *Perspectives on Psychological Science, 3,* 65–72.

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

Haidt, J., Björklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. *Unpublished manuscript, University of Virginia*.

Haidt, J. & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20,* 98–116.

Haidt, J. & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology, 31,* 191–221.

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

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.

Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. C. (2009). Disgust and the moralization of purity. *Journal of personality and social psychology, 97,* 963–976.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences, 13,* 1–6.

Huebner, B., Lee, J. J., & Hauser, M. D. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture, 10,* 1–26.

Inbar, Y., Pizarro, D. A., & Bloom, P. (2009). Conservatives are more easily disgusted than liberals. *Cognition and Emotion, 23,* 714–725.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making, 8,* 407–424.

Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect, and the moral/conventional distinction. *Mind and Language, 22,* 117–131.

Landy, J. F. & Bartels, D. M. (2016). *The moral violations database: A large, normed set of behavioral descriptions.* Manuscript in preparation.

Markovits, H. & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical syllogisms. *Memory and Cognition, 17,* 11–17.

Medin, D. L. & Rips, L. J. (2005). Concepts and categories: Memory, meaning, and metaphysics. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 37–72). Cambridge, UK: Cambridge University Press.

Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition and Emotion, 16,* 695–703.

Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development, 52,* 114–121.

Nucci, L. P. & Nucci, M. S. (1982). Children's responses to moral and social conventional transgressions in free-play settings. *Child Development, 53,* 1337–1342.

Nucci, L. P. & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development, 49,* 400–407.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). The role of analytic thinking in moral judgments and values. *Thinking and Reasoning, 20,* 188–214.

Royzman, E. B., Atanasov, P., Landy, J. F., Parks, A., & Gepty, A. (2014). CAD or MAD? Anger (not disgust) as the predominant response to pathogen-free violations of the Divinity code. *Emotion, 14,* 892–907.

Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making, 10,* 296–313.

Royzman, E. B., Landy, J. F., & Goodwin, G. P. (2014). Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults. *Judgment and Decision Making, 9,* 176–190.

Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral–conventional distinction. *Cognition, 112,* 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,* 574–586.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47,* 609–612.

Shweder, R. A. & Much, N. C. (1991). Determinations of meaning: Discourse and moral socialization. In. R. A. Shweder, *Thinking through cultures: Expeditions in in cultural psychology* (pp. 186–240). Cambridge, MA: Harvard University Press.

Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook*

of moral development (pp. 23–45). New York: NY: Psychology Press.

Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., & Comer, J. (2012). Developmental changes and individual differences in young children's moral judgments. *Child Development, 83*, 683–696.

Sousa, P. (2009). On testing the "moral law". *Mind and Language, 24,* 209–234.

Sousa, P., Holbrook, C., & Piazza, J. (2009). The morality of harm. *Cognition, 113,* 80–92.

Stitch, S., Fessler, D. M. T., & Kelly, D. (2009). On the morality of harm: A response to Sousa, Holbrook and Piazza. *Cognition, 113*, 93–97.

Talhelm, T., Haidt, J., Oishi, S., Zhang, X., Miao, F. F., & Chen, S. (2015). Liberals think more analytically (more "WEIRD") than conservatives. *Personality and Social Psychology Bulletin, 41,* 250–267.

Tisak, M. S. & Turiel, E. (1984). Children's conceptions of moral and prudential rules. *Child Development, 55,* 1030–1039.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* Cambridge, UK: Cambridge University Press.

Turiel, E. (2002). *The culture of morality.* Cambridge, UK: Cambridge University Press.

Turiel, E. (2014). Morality: Epistemology, development, and social opposition. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 3–22). New York: NY: Psychology Press.