Mental states are more important in evaluating moral than conventional violations

Carly Giffin (carly.giffin@berkeley.edu)
Tania Lombrozo (lombrozo@berkeley.edu)
Department of Psychology, 3210 Tolman Hall, Berkeley, CA 94720-1650 USA

Abstract

A perpetrator's mental state – whether she had *mens rea* or a "guilty mind" – typically plays an important role in evaluating wrongness and assigning punishment. In two experiments, we find that this role for mental states is weaker in evaluating conventional violations relative to moral violations. We also find that this diminished role for mental states may be associated with the fact that conventional violations are wrong by virtue of having violated a (potentially arbitrary) rule, whereas moral violations are also wrong inherently.

Keywords: decision making, violations, mental states, moral evaluation, punishment.

Introduction

Both folk intuitions and the law accord a prominent role to mental states when it comes to assessing the severity of a transgression and how it should be punished. For example, serving someone a cup of coffee sprinkled with poison is deemed quite a bit worse when it was done intentionally – with full knowledge that the coffee contained poison – than when it resulted from the false belief that the poison was sugar (Young et al., 2007). To take a legal example, determinations of whether a defendant should be sentenced with murder versus manslaughter depend, in large part, on whether the killing was intentional.

Nonetheless, recent findings point to the idea that mental states are not equally important for all types of transgressions. Young and Saxe (2011), for example, find that an offender's knowledge has a greater impact on how people evaluate a harm violation as opposed to a purity violation (see also Hawley-Dolan & Young, 2013; Russell & Giner-Sorolla, 2011). In legal judgments, we have found that mental states play a weaker role in judgments concerning *strict liability* crimes, such as speeding or statutory rape, relative to crimes that are not strict liability, such as burglary or battery, for which the presence of *mens rea*, a "guilty mind," informs a defendant's conviction and sentence (Giffin & Lombrozo, in prep). In other words, transgressions seem to differ in the extent to which they are moderated by the perpetrator's mental states, such as her beliefs and intentions—a property which we refer to as "knowledge dependence."

Why might this be? Here, again, a legal distinction is useful: that between transgressions that are *malum in se*, or wrong in themselves, versus *malum prohibitum*, or wrong because they are prohibited (*US v. Morissette*, 1952). Consider, for example, a moral violation, such as murder. Even in the absence of a rule prohibiting the action, we would consider it morally wrong. Violations of convention, in contrast, are problematic *because* they violate the convention: there's nothing inherently right or wrong about going 50 miles per hour, unless you're in a 35 mile per hour zone. In light of the rule and its consequences for others, however, the action becomes problematic.

The distinction between moral and conventional violations is familiar from research in moral psychology as well. In classic experiments, Turiel and colleagues presented children with stories in which an actor violated a rule, and were asked to judge how bad the actor's behavior was, both with the rule in place and in a situation in which it didn't apply (Turiel, 2008; Weston & Turiel, 1980). They found that children as young as six judged conventional violations (such as violating a dress code) but not moral violations (such as hitting another child) in a way that was highly "rule dependent": in the absence of a dress code the child's dress was just fine, but in the absence of a rule about hitting, hitting another child was still wrong.

These judgments concerning the wrongness of an action presumably stem, in part, from participants' assessments of the "wrongness" of the perpetrator's beliefs and intentions. Knowingly hitting a child is wrong regardless of the rules; one should never intend to hit a peer. Knowingly wearing a t-shirt, however, is only wrong in certain conditions; the intention to wear a t-shirt and the belief that one is doing so are not, on their own, problematic. To the extent that evaluations of wrongness and punishment stem from evaluations of the underlying mental states, and not just from the violation of a rule, one might expect moral violations to be more knowledge dependent than conventional violations. In two experiments, we test this prediction.

Experiment 1

In Experiment 1, we test the prediction that mental states play a larger role in the evaluation of violations of moral rules (hereafter referred to as "moral violations") relative to violations of conventional rules (hereafter referred to as "conventional violations"). To do so, we compare judgments of wrongness and punishment across stories involving moral or conventional transgressions of a stipulated rule, where the violation is committed *knowingly* or *unknowingly* (i.e., due to an accident or false belief concerning something other than the rule itself). In other words, we test the prediction that the evaluation of moral violations is more "knowledge dependent" than the evaluation of conventional violations. We also replicate the well-established finding that judgments concerning conventional violations tend to be more "rule dependent"

than those concerning moral violations, and investigate the relationship between knowledge dependence and rule dependence by testing whether violations that generated greater knowledge effects tended to generate weaker effects of a rule change.

Methods

Participants. One-hundred-and-sixty adults (96 female, 64 male, mean age = 36, *SD* = 12) participated in the study through Amazon Mechanical Turk in exchange for monetary compensation. An additional 69 participants were tested but excluded for failing catch questions (55) or to ensure even numbers in all conditions (14). Participation was restricted to workers with IP addresses in the United States and a prior HIT approval rating of 95% or higher.

Materials & Procedure. The experimental stimuli consisted of 12 distinct stories, 6 of which involved conventional violations and 6 of which involved moral violations. There were two versions of each story: one involved an agent who committed the violation knowingly, and one an agent who knew the rule, but violated it unknowingly.

Six of the stories (Teacher's Title, Greeting, Baseball, Dollar, Physician, and Embezzler) were based on vignettes originally presented to children by Davidson, Turiel, and Black (1983). These stories were modified to generate matched knowing and unknowing conditions. The Physician and Embezzler vignettes were additionally modified to take place in a school setting.

Individual participants were randomly assigned to one of four conditions, the result of crossing violation domain (2: conventional, moral) with knowledge status (2: knowing, unknowing). Each participant received the corresponding six vignettes in a random order. Sample *knowing* and *unknowing* vignettes for one story, Baseball, are excerpted below. In this story, the rule was that students had to wear a blue shirt with the school logo on the back to practice:

> *Knowing*: "One day, Jack was getting ready for a baseball practice. He was tired of always wearing his blue practice shirt; he thought it would be fun to wear another shirt for a change. So Jack went to the practice wearing a blue shirt that did not have the school logo on the back."
>
> *Unknowing:* "One day, Jack was getting ready for a baseball practice. He was in a hurry to get to the bus on time, so he dressed quickly and left. Jack didn't realize he had grabbed the wrong blue shirt. So Jack went to the practice wearing a blue shirt that did not have the school logo on the back."

The presentation of each vignette was first followed by censure and detention questions, presented on one screen in random order:

> *Censure*: "How wrong was [Actor's actions]?" Participants indicated their answer on a scale from 0 (not at all wrong) to 6 (very wrong).
>
> *Detention*: "How many hours of detention should [Actor] get?" Participants indicated their answer on a scale from 0 to 6 hours.

After answering these two questions, participants were presented with another screen and asked to indicate the censure and detention ratings the actor would deserve if the school had never had a rule prohibiting the action. The wording of these questions (again presented in random order) was identical to those above, but preceded by the following: "What if [Actor's] school had no rule prohibiting what [s]he did? Please answer the following questions based on this rule change."

Next, on a separate screen, participants were presented with a true/false question relating to the vignette they had just read. These questions were used to assess whether the participants had read the vignette carefully; those who answered any comprehension questions incorrectly were excluded from further analyses.

After reading all six vignettes and answering their associated questions, participants answered one additional catch question designed to ensure that they were reading instructions carefully, modeled after Oppenheimer, Meyvis, and Davidenko (2009). Finally, participants answered demographic questions about their age and gender.

Results

Initial censure and detention ratings. To test the prediction that judgments regarding conventional violations are less "knowledge dependent" than those regarding moral violations, we performed a 2 (knowledge status: knowing, unknowing) x 2 (violation domain: conventional, moral) ANOVA on initial censure ratings, and another on initial detention ratings (see Figure 1). We expected to find an interaction between knowledge status and violation domain, with a larger effect of knowledge status for moral violations than for conventional violations.

Both analyses revealed main effects of violation domain. Moral crimes received significantly higher initial censure, $F(1,156) = 87.72$, $p<.000$, $\eta_p^2 = .36$, and detention ratings, $F(1,156) = 169.25$, $p<.000$, $\eta_p^2 = .52$. The analyses also revealed main effects of knowledge status, with censure, $F(1,156) = 63.13$, $p<.000$, $\eta_p^2 = .29$, and detention, $F(1,156) = 40.5$, $p<.000$, $\eta_p^2 = .21$, ratings significantly higher for the knowing condition than the unknowing condition.

Most critically, we found the predicted interaction for both censure, $F(1,156) = 19.02$, $p<.000$, $\eta_p^2 = .11$, and detention, $F(1,156) = 24.00$, $p<.000$ $\eta_p^2 = .13$. Independent samples t-tests showed that censure ratings were significantly higher for the knowing condition than for the unknowing condition for both conventional violations, $t(78)$

= 2.6 (p<.012), d = .57, and for moral violations, t(78) = 8.62 (p<.001), d = 1.95, but the effect was greater for moral violations. Similarly, for detention ratings, responses were higher for the knowing condition than for the unknowing condition for both conventional violations, t(78) = 1.99 (p<.050), d = .45, and moral violations, t(78) = 6.06 (p<.001), d = .1.37, but the effect was greater for moral violations.

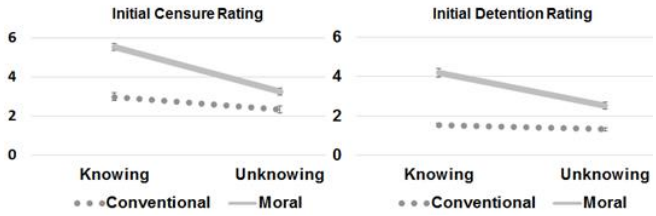

Figure 1: Interaction between knowledge condition and violation type for initial censure and detention ratings. Error bars correspond to one SEM in each direction.

Censure and detention ratings: effects of rule revocation. To what extent are conventional and moral violations condemned *because* they violate a rule – that is, to what extent are judgments of wrongness and punishment "rule dependent" in each domain? To address these questions, censure and detention difference scores were created by subtracting participants' censure and detention ratings *after* the rule change from their corresponding initial scores. We performed 2 (knowledge status: knowing, unknowing) x 2 (violation domain: conventional, moral) ANOVAs on each difference score, and predicted a greater drop in both ratings for conventional violations relative to moral violations.

This prediction was confirmed for censure: we found a significant main effect of violation domain, F(1,156) = 33.76, p<.000, η_p^2 =.18, with a larger drop in the perceived wrongness of conventional violations following the revocation relative to moral violations (see Figure 2). The ratings for detention, however, were unexpected: we found a significant main effect of domain, but in the opposite direction, F(1,156) = 5.37, p<.022, η_p^2 =.03. This may be in part because conventional violations were assigned very low levels of punishment; the response distribution was skewed, with little room for a drop in ratings.

We also found a significant main effect of knowledge for censure ratings, with a larger drop in perceived wrongness for the knowledge condition (M=.91, .STD=94) relative to the false belief conditions (M=.61, STD=.72), F(1,156) = 6.45, p<.012, η_p^2 =.04. There were no additional significant effects.

Analysis across violations: rule dependence and knowledge dependence. To investigate the relationship between knowledge dependence and rule dependence in a more fine-grained way, we examined whether those stories that were the most knowledge dependent were also the least rule dependent.

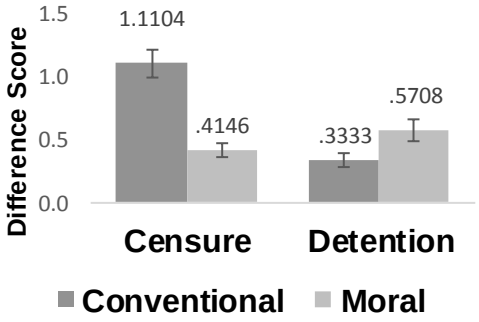

Figure 2: Rule dependence for censure and detention ratings (ratings after rule change subtracted from initial ratings). Error bars correspond to one SEM in each direction.

To obtain a measure of knowledge dependence for each story, the average censure ratings from participants in the unknowing condition for that story were subtracted from the average censure ratings from participants in the knowing condition for that story. To obtain a measure of rule dependence for each story, we averaged the censure difference scores (initial censure ratings minus censure ratings after rule change) for the knowledge and false belief versions of that story. This resulted in twelve pairs of numbers – one pair for each story – that revealed a significant negative correlation, r = -.74, p < .006, as predicted (see Figure 3): those stories for which wrongness judgments were more knowledge dependent tended to involve judgments that were less rule dependent.

Considering just conventional violations, the correlations between rule dependence and knowledge dependence for both censure and detention were positive, but they were

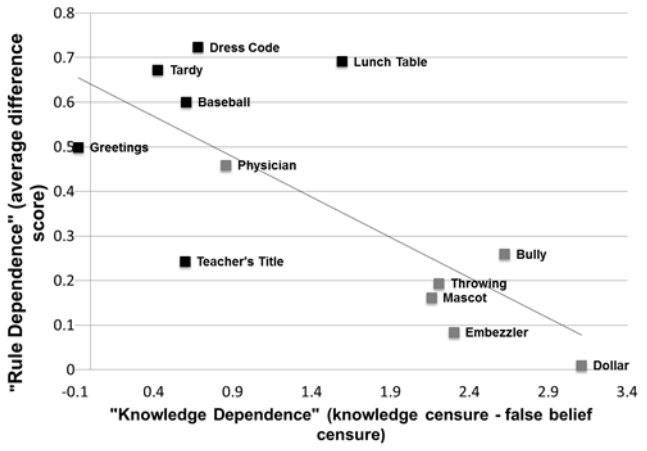

Figure 3: Significant negative correlation, across stories, between rule dependence and knowledge dependence. Moral stories are indicated with black; conventional stories with dark grey.

not significant. For moral violations, both the censure, $r = -.86$, $p < .027$, and detention, $r = -.88$, $p < .019$, correlations were negative and significant. Comparable analyses with detention ratings did not yield any significant effects.

Discussion

Experiment 1 found a novel relationship between violation type (moral versus conventional) and the extent to which mental states – in this case, acting knowingly or unknowingly – influence moral judgments. Specifically, we found that participants judged actors more harshly (in terms of both the wrongness of their action and the hours of detention deserved) when they violated a rule knowingly as opposed to unknowingly. However, the effect of knowledge status was significantly greater for moral violations relative to conventional violations.

It is worth emphasizing that the knowledge that was varied across conditions was not knowledge of the rule (e.g., how one should dress for baseball practice). Instead, the characters who committed violations unknowingly did so because they had a false belief that affected whether an action counted as a transgression with respect to the known rule. For example, in the Baseball story, the character knew that he was supposed to wear a blue shirt with the school logo on the back to practice, but had a false belief that the blue shirt he had put on in fact had the logo on the back.

Consistent with prior work by Turiel and colleagues, Experiment 1 also found that when the operating rule was revoked, judgments of wrongness were reduced to a greater extent for conventional violations than for moral violations. Puzzlingly, the same pattern did not emerge for judgments concerning punishment (hours of detention). This may be because detention ratings were generally low, especially for conventional violations – participants may not have felt that detention was an appropriate punishment to be considering, and there was little room for a change in ratings across conditions.

Finally, Experiment 1 provides hints that "knowledge dependence" – which is stronger for moral than conventional violations – is negatively associated with "rule dependence" – which is stronger for conventional than moral violations. In an analysis across stories, we found that those stories associated with the greatest knowledge dependence tended to show greater rule dependence. However, this effect was driven entirely by the moral stories.

Experiment 2

Experiment 2 had three aims. First, we sought to replicate the novel finding from Experiment 1 that knowledge status has a larger impact in evaluations of moral violations relative to conventional violations, and the more familiar finding that rule revocation has a larger impact on the perceived wrongness of a conventional violation than of a moral violation. Second, we hoped to clarify the relationship between rule revocation and punishment by considering an alternative to detention: school service hours. We expected that school service hours would be deemed a more appropriate punishment for the violations considered, thereby avoiding a floor effect on ratings. Finally, Experiment 2 manipulated *both* knowledge status and rule status within subjects, thus allowing us to evaluate the relationship between knowledge dependence and rule dependence across participants, not only across stories.

Methods

Participants. Two-hundred-and-forty adults (114 female, 152 male, 2 other/prefer not to specify, mean age = 32, *SD* = 14) participated in the study through Amazon Mechanical Turk in exchange for monetary compensation. An additional 28 participants were tested, but were excluded for failing catch questions (27) or to ensure even numbers in all conditions (1). Participation was restricted to workers with IP addresses in the United States and with a prior HIT approval rating of 95% or higher.

Materials & Procedure. As in Experiment 1, the experimental stimuli consisted of 12 distinct stories, 6 of which concerned conventional violations and 6 of which concerned moral violations, leading to 12 conditions. Eleven stories were similar to those from Experiment 1, with the Physician Story replaced entirely by the Pushing story, driven by concerns that the Physician story involved both conventional and moral aspects that made it a poor representative of a moral violation. In the Pushing story, adapted from Davidson, Turiel, and Black (1983), an actor either knowingly or unknowingly pushes another student down. Several of the conventional stories from Experiment 1 were also modified in an attempt to increase their perceived wrongness and therefore make them more comparable in severity to their moral counterparts.

Each participant read only one story. Participants first read the false belief version of their assigned story and answered two evaluative questions. The censure question was identical to that in Experiment 1; the punishment question was changed to the following:

> *School Service*. Students who break a rule at [Actor's] school are given school service hours during which they clean classrooms, organize supplies, and pick up trash on the grounds. How many hours of school improvement service should [Actor] get?

Participants were then asked to imagine that the actor had instead violated the rule knowingly, and again rated wrongness and punishment. Below is a sample from the Baseball story:

> *Knowledge Change.* "Suppose that Jack had actually realized, while he was dressing, that the shirt he was about to put on for practice violated the rule – that is, that it didn't have the logo on the back. And suppose that he decided to wear it anyway. In this case, where Jack

knowingly violated the rule, how would you respond to the following questions? (Your responses may be the same as those you just provided, or they may differ.)"

Finally, as in Experiment 1, participants were told to imagine that the rule had been revoked, and answered the evaluative questions a final time. Below is a sample from the Baseball story:

Rule Change. "Finally, suppose that Jack's school had no rule prohibiting wearing a shirt without the school logo to practice, and Jack knowingly wore a shirt without the school logo to practice. In this case, with no rule about how to dress for practice in place, how would you respond to the following questions? (Your responses may be the same as those you've provided, or they may differ.)"

After reading all versions of the story and answering their associated questions, participants answered the same catch and demographic questions as in Experiment 1.

Results

Effects of knowledge on censure and punishment. To create a measure of knowledge dependence for each dependent variable, we subtracted the first censure and service hour ratings, taken after reading the unknowing story, from the second set of ratings, corresponding to the knowing violation. Independent t-tests were performed on both difference scores, and we predicted a greater difference for moral relative to conventional violations.

As predicted, these analyses found that the knowledge effect was greater for moral than conventional violations for both censure, $t(238)= 4.12$, $p< .001$, $d = .53$, and service hours, $t(221)= 5.97$, $p< .001$, $d = .80$ (corrected for violating Levene's; See Figure 4).

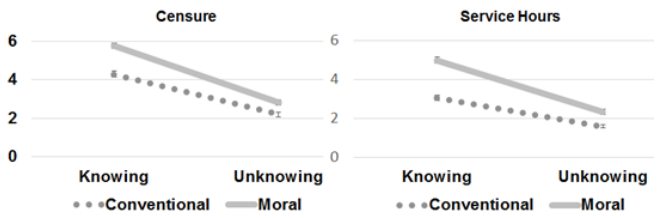

Figure 4: Ratings for censure and service hours for both the knowing and unknowing conditions as a function of violation type. Error bars correspond to one SEM in each direction.

Effect of rules on censure and punishment. The contribution of the violation of a rule ("rule dependence") was measured by subtracting participants' censure and detention ratings after the rule change from their corresponding scores from the initial vignettes, averaged across the knowing and unknowing conditions. Independent *t*-tests were performed on these difference scores, and we predicted a greater difference for conventional relative to moral violations.

As predicted, these analysis found that rule dependence was significantly greater for conventional than moral violations for both censure, $t(220)= 9.42$, $p < .001$, $d = 1.27$ and service hours, $t(181)= 5.00$, $p < .001$, $d= .74$, (both corrected for violating Levene's). (See Figure 5).

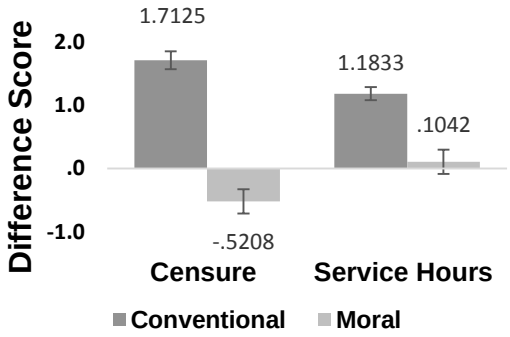

Figure 5: Rule dependence of censure and service hours (i.e., ratings after rule change subtracted from average of initial ratings). Error bars correspond to one SEM in each direction.

Relationship between rule dependence and knowledge dependence. In Experiment 1, we investigated the relationship between knowledge dependence and rule dependence across stories, i.e., by analyzing whether those stories that generated the most knowledge-dependent judgments also generated the least rule-dependent judgments. In the present case, methodological changes allowed us to create a measure of rule dependence and knowledge dependence for each participant, allowing us to test for a relationship across *participants* rather than only across *stories*. We predicted a negative correlation, with participants who were more influenced by mental states being less influenced by the rule change.

We ran bivariate correlations between the difference scores reflecting knowledge dependence and those reflecting rule dependence. Consistent with our prediction, these scores were negatively correlated for censure, $r = -.13$, $p < .039$; however, the correlation for service hours was not significant, $r = -.005$. $p < .94$. Analyzing conventional violations only, the correlation was *positive* and significant for censure, $r = .32$, $p < .001$, and service hours, $r = .70$, $p < .001$. Analyzing moral violations only, the correlations for both censure, $r = -.22$, $p <. 014$, and service hours, $r = -.11$, $p < .241$, were negative, but the latter was not significant.

Discussion

Experiment 2 successfully replicated key findings from Experiment 1. Specifically, we found (once again) that relative to judgments concerning conventional violations, those concerning moral violations were more sensitive to whether the actor transgressed knowingly versus unknowingly. We also found that relative to judgments concerning moral violations, those concerning conventional violations were more influenced by a rule revocation. This was the case both for judgments of wrongness and for our new measure of punishment, which involved service hours as opposed to detention. Finally, we also found the predicted negative relationship between rule dependence and knowledge dependence, but with an analysis that considered variation across participants rather than only across stories. This relationship was significant for judgments of wrongness. However, when each type of violation was considered in isolation, it became clear that the negative relationship was driven by moral violations, as in Experiment 1. Conventional violations actually showed a positive relationship between knowledge dependence and rule dependence.

General Discussion

In two experiments we find evidence that the evaluation of moral violations is more knowledge dependent than the evaluation of conventional violations, while the evaluation of conventional violations is more rule dependent than that of moral violations. We also find evidence that these properties are negatively associated for moral violations, whether the analysis is across violations (Experiment 1) or across participants (Experiment 2).

The finding that conventional violations are more rule dependent is not new; however, to our knowledge, this is the first demonstration that conventional violations are also less sensitive to mental states. This relationship between knowledge dependence and the conventionality of a rule is consistent with prior work on the evaluation of strict liability crimes (Giffin & Lombrozo, in prep), where we argue that knowledge is less important in folk judgments concerning strict liability crimes because such crimes tend to involve the violation of a rule with somewhat arbitrary – and therefore conventional – elements. For example, speeding involves the violation of a somewhat arbitrary speed limit. Driving 40 miles per hour is not inherently wrong, but it is wrong when it occurs in a 35-mile zone, and the designation of 35 miles (as opposed to 34 or 36.5) as the limit is somewhat arbitrary. This feature of "arbitrariness" could potentially help explain why strict liability crimes behave more like conventional violations. This feature may also relate to why mental states play a weaker role, since knowing that one is engaging in a particular act (e.g., driving 40 miles per hour), even when doing so knowingly, is not itself inherently wrong.

While our predictions held robustly for judgments of "wrongness," the findings concerning punishment (detention and service hours) were more mixed. Prior work suggests that mental states may be more important when evaluating moral wrongness than when ascribing punishment (Cushman, 2008), which could help explain this dissociation. Our finding of a negative association between knowledge dependence and rule dependence was also inconsistent across cases: while it held for moral violations in both Experiment 1 and Experiment 2, conventional violations not only failed to exhibit this relationship, but had a positive relationship (not significantly in Experiment 1, and significantly in Experiment 2). Why this is so is an important question for future research.

In sum, our findings are consistent with prior work demonstrating the importance of mental states in moral judgment, and establish a previously-undocumented relationship between knowledge dependence and rule dependence. This raises important questions about precisely which mental states matter in evaluating different transgressions and why.

References

Cushman, F. (2008). Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment. *Cognition, 108*, 353-380.

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.

Giffin, C., & Lombrozo, T. (2014). Wrong or Merely Prohibited: Special Treatment of Strict Liability Crimes in Folk Judgment. Manuscript in preparation.

Russell, P., & Giner-Sorolla, R. (2011). Moral Anger, but Not moral Disgust, Responds to Intentionality. *Emotion, 11*(2), 233-240.

Turiel, E. (2008). Thought About Actions in Social Domains: Morality, Social Conventions, and Social Interactions. (2008). *Cognitive Development, 23*, 136-154.

United States v. Morissette, 342 U.S. 246 (1952).

Weston, D. & Turiel, E. (1980). Act-Rule Relations: Children's Concepts of Social Rules. *Developmental Psychology, 16(5)*, 417-424.

Young, L., Cushman, F., Hauser, M., Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *PNAS,104* (20), 8235-8240. doi:10.1073/pnas.0701408104

Young, L., & Saxe, R. (2011). When Ignorance is No Excuse: Different Roles for Intent Across Moral Domains. *Cognition, 3*, 202–214.

Acknowledgements

We gratefully acknowledge Elliot Turiel for helpful discussion and for sharing stimulus materials. This work was partially supported by a McDonnell Foundation Scholar Award to the second author.