Cognitive Science 33 (2009) 273–286

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01013.x

The Role of Moral Commitments in Moral Judgment

Tania Lombrozo

Department of Psychology, University of California, Berkeley

Received 25 February 2008; received in revised form 3 June 2008; accepted 9 July 2008

Abstract

Traditional approaches to moral psychology assumed that moral judgments resulted from the application of explicit commitments, such as those embodied in consequentialist or deontological philosophies. In contrast, recent work suggests that moral judgments often result from unconscious or emotional processes, with explicit commitments generated post hoc. This paper explores the intermediate position that moral commitments mediate moral judgments, but not through their explicit and consistent application in the course of judgment. An experiment with 336 participants finds that individuals vary in the extent to which their moral commitments are consequentialist or deontological, and that this variation is systematically but imperfectly related to the moral judgments elicited by trolley car problems. Consequentialist participants find action in trolley car scenarios more permissible than do deontologists, and only consequentialists moderate their judgments when scenarios that typically elicit different intuitions are presented side by side. The findings emphasize the need for a theory of moral reasoning that can accommodate both the associations and dissociations between moral commitments and moral judgments.

Keywords: Moral judgment; Moral reasoning; Moral intuition; Moral dilemmas; Trolley car problems; Consequentialism; Deontology

> *My work is based on the assumption that clarity and consistency in our moral thinking is likely, in the long run, to lead us to hold better views on ethical issues.*
>
> – Peter Singer (Singer, 2003, pg. 53)

Individuals vary in their moral commitments concerning torture, the treatment of nonhuman animals, and a host of equally controversial issues, making moral disagreements a source of heated contemporary debate. Attempting to resolve such disagreements through explicit reasoning and argumentation reveals two assumptions behind

Correspondence should be sent to Tania Lombrozo, Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720. E-mail: lombrozo@berkeley.edu

the very idea of progress when it comes to the domain of morality. The first is that individual variation in moral judgments (e.g., about the moral status of abortion, the death penalty, or meat consumption) is mirrored by variation in explicitly held moral commitments (e.g., about the conditions under which ending a life is permissible). The second is that changing explicitly held moral commitments—the typical aim of argumentation—can impact moral judgments, and thus that moral judgments are generated from moral commitments.

These assumptions resonate with traditional approaches to moral judgment within psychology (e.g., Kohlberg, 1969; Turiel, 1983). Kohlberg, for example, studied explicit moral reasoning by eliciting justifications for moral judgments, which suggests such reasoning is the basis for judgments (Kohlberg, 1969). However, more recent approaches to moral cognition challenge assumptions about the coherence of explicitly held moral commitments and about their role in generating moral judgments. For example, a phenomenon called ''moral dumbfounding'' reveals that many moral judgments are not accompanied by adequate justifications (Haidt, 2001). This suggests that the moral commitments identified in justifications are not causally responsible for the corresponding judgments (Haidt, 2001; Hauser, Cushman, Young, Jin, & Mikhail, 2007). The dissociation between moral judgments and justifications has led a few prominent theorists to posit two distinct systems that play a role in moral evaluation: one that features immediate and emotion-driven intuitions, the other deliberate and logical reasoning (e.g., Greene, 2007; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Somerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001).

The contrast between traditional and more contemporary approaches to moral psychology can be illustrated by a class of moral dilemmas known as trolley car problems (e.g., Cushman, Young, & Hauser, 2006; Foot, 1967; Greene et al., 2001; Hauser, 2006; Mikhail, 2007; Petrinovich, O'Neill, & Jorgensen, 1993; Thompson, 1985). In one scenario, ''switch,'' a bystander can redirect a trolley about to kill five people to a track containing only one person. In another scenario, ''push,'' a bystander can halt the same trolley by pushing a large man into its path, killing the one to save five. Both scenarios involve a trade-off between one life and five, but they differ in the means involved. Studies find that a majority of participants find action in scenarios like ''switch'' permissible and action in scenarios like ''push'' impermissible, and that this response pattern is driven by a number of factors (e.g., Cushman et al., 2006; Greene et al., 2001; Hauser et al., 2007; Mikhail, 2007; Waldmann & Dieterich, 2007).

Trolley car scenarios reveal the opposition between two families of moral philosophies: those that evaluate the permissibility of actions in terms of actual or expected consequences, called consequentialist, and those that evaluate the permissibility of actions in terms of rules, rights, or obligations, called deontological. A consequentialist should treat the ''switch'' and ''push'' scenarios equivalently and find action in each case permissible: one death is a better consequence than five.[1] A deontologist might find action impermissible and distinguish the two scenarios, as they differ in means if not in consequences. In particular, a deontologist can invoke the doctrine of double effect, the principle that it is permissible to cause harm as a foreseen side effect of a greater good (as in the ''switch''

scenario), but not as a means to a greater good (as in the ''push'' scenario) (e.g., McIntyre, 2006).

A traditional approach to moral psychology would assume that judgments result from the application of moral commitments, such as an explicit endorsement of the doctrine of double effect. While a principle like the doctrine of double effect does account for variation in judgments across scenarios (Cushman et al., 2006; Mikhail, 2007), participants almost never invoke this principle when asked to justify their judgments (Cushman et al., 2006; Hauser et al., 2007). This finding reaffirms the view from contemporary moral psychology that explicitly accessible moral commitments of the kind that figure in justifications and argumentation are not the only, or even the primary, basis for moral judgment. Instead, unconscious principles (Hauser, 2006; Mikhail, 2007) or immediate, emotional reactions (Greene et al., 2001; Haidt, 2001) may predominate.

The current paper explores a middle course between the traditional view that moral commitments generate moral judgments and a more contemporary view that moral commitments and moral judgments are largely independent. The motivating hypothesis is that moral commitments and moral judgments are systematically related, but that this systematic relationship need not result from the explicit application of moral commitments in the course of moral judgment. Rather, moral commitments may causally mediate moral judgments, but gradually or indirectly, perhaps by changing the features of scenarios that are judged morally relevant or that elicit immediate intuitions.

In the experiment that follows, participants completed a task designed to assess whether their explicit moral commitments were more deontological or consequentialist and a second task involving trolley car problems. The hypothesis that moral commitments and moral judgments are systematically related predicts that variation in trolley car judgments should track variation in explicit moral commitments, even if the relationship is imperfect. To examine the stronger claim that moral commitments partially *mediate* responses on trolley car problems, the experiment involves an additional manipulation: whether the ''switch'' and ''push'' scenarios are evaluated in isolation, as in a between-subjects design, or together, as in a within-subjects design. Previous work suggests that evaluating scenarios together can facilitate the application of rules (Gentner & Medina, 1998), increase conformity to normative beliefs (Bazerman, Tenbrunsel, & Wade-Benzoni, 1998; O'Connor et al., 2002), and promote more reflective and comparative thought (Bazerman, Moore, Tenbrunsel, Wade-Benzoni, & Blount, 1999; Hsee, Loewenstein, Blount, & Bazerman, 1999; see also Nichols & Knobe, 2007). If moral commitments mediate judgments on trolley car problems, then seeing the scenarios together should facilitate the application of individuals' commitments and thus exaggerate differences between deontological and consequentialist participants. Alternatively, if differences between deontological and consequentialist responding reflect differences in reflective thought or reliance on emotion, as some accounts would suggest (e.g., Greene, 2007; Greene et al., 2008), then seeing the scenarios together might lead all participants toward more consequentialist responses.

1. Methods

1.1. Participants

Three hundred thirty-six participants completed the study (112 male, 224 female, mean age = 20 years). The majority were Berkeley undergraduates who participated for course credit; a minority were summer school students and volunteers recruited from the campus community.

1.2. Materials and procedures

Participants completed a questionnaire consisting of two tasks presented in counterbalanced order and separated by an unrelated filler task. In the *Commitments* task, participants responded to six questions of the following form (labels in italics for reference):

> Which of the following statements best characterizes your position on lying?
>
> *Deontological response*: It is never morally permissible to lie.
>
> *Consequentialist response*: If lying will produce greater net good than bad, then it is morally permissible to lie.
>
> *Strong consequentialist response*: If lying will produce greater net good than bad, then it is morally obligatory to lie.

The other five questions concerned assassination, torture, murder, stealing, and forced sterilization. Participants were explicitly told to respond on the basis of their moral convictions, without regard for whether an action is legal. The six questions were presented in one of four orders, and question order was counterbalanced with the other experimental variables.

In the *Judgments* task, participants were presented with trolley car problems modified from Hauser et al. (2007). One third of participants were randomly assigned to the *Switch* condition, in which they read the following scenario and made the judgments that follow (only the permissibility judgments will be discussed):

> David is a passenger on a train. The driver just shouted that the train's brakes have failed and has fainted out of shock. On the track ahead are five people; the banks are so steep that they will not be able to get off the track in time. The track has a side track leading off to the left, and David can turn the train onto it. Unfortunately there is one person on the side track. David can turn the train, killing the one; or he can refrain from turning the train, letting the five die.
>
> Is it morally permissible for David to switch the train to the side track?
> 1 2 3 4 5 6
> *Definitely not*……*Definitely*

If David switches the train to the side track, should he be punished?
1 2 3 4 5 6
Definitely not……*Definitely*

If David fails to switch the train to the side track, should he be punished?
1 2 3 4 5 6
Definitely not……*Definitely*

One third of participants were randomly assigned to the *Push* condition, in which they read the following scenario and made the judgments that follow.

Fred is on a footbridge over the train tracks. He knows trains and can see that the one approaching the bridge is out of control. On the track under the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. Fred knows that the only way to stop an out-of-control train is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a large man wearing a backpack, also watching the train from the footbridge. Fred can shove the man with the backpack onto the track in the path of the train, killing him; or he can refrain from doing this, letting the five die.

Is it morally permissible for Fred to shove the man?
1 2 3 4 5 6
Definitely not……*Definitely*

If Fred shoves the man, should he be punished?
1 2 3 4 5 6
Definitely not……*Definitely*

If Fred fails to shove the man, should he be punished?
1 2 3 4 5 6
Definitely not……*Definitely*

The remaining third of participants were in the *Both* condition, in which they read both the switch and push scenarios, and only after reading both were they asked to make the corresponding judgments for each scenario. The scenario order in this condition was counterbalanced.

2. Results

Although the experimental hypotheses concern the relationship between responses on the *Commitments* task and on the *Judgments* task, I present the data from each task individually before considering interactions. All reported means are followed in parentheses by the

standard deviation, and *p*-values are for two-tailed tests. Homogeneity of variance was verified with Levene's test for equality of variances, and it did not differ across groups except as noted.

2.1. Data on moral commitments

Fig. 1 illustrates the proportion of choices for each of the six queried actions. Responses did not vary as a function of question order [F(3,332) = .01–.36, *n.s.*] nor as a function of task order [$F(1,334) = .08–.86$, *n.s.*]. The internal consistency of the items was moderate (Cronbach's alpha = .65). An aggregate score was created by treating a deontological response as ''0,'' a consequentialist response as ''1,'' a strong consequentialist response as ''2,'' and averaging the six responses for each participant. The resulting score, which I will refer to as a consequentialism score, reflects the extent to which an individual provides consequentialist responses. The population mean for consequentialism score was .52 (.32) (see Fig. 2). There was a small but significant effect of sex [$t(183) = 3.03$, $p < .01$, equal variances not assumed, $r = .22$], with men generating more consequentialist responses than women, .59 (.36) vs. .48 (.28), and also having significantly greater variance in scores than women [Levene's test for equality of variances, $F(1,334) = 10.41$, $p < .01$]. This sex difference was largely driven by three items that yielded significant sex differences on post hoc tests: assassination, .65 (.63) vs. .48 (.57) [$t(334) = 2.56$, $p < .05$, $r = .14$], torture, .41 (.59) vs. .26 (.46) [$t(180) = 2.30$, $p < .05$, equal variances not assumed, $r = .17$], and murder, .45 (.55) vs. .29 (.47) [$t(192) = 2.50$, $p < .05$, equal variances not assumed, $r = .18$].

2.2. Data on moral judgments

Replicating past work, participants rated action in the switch scenario more permissible than in the push scenario (see Fig. 3), whether the judgments were made between subjects

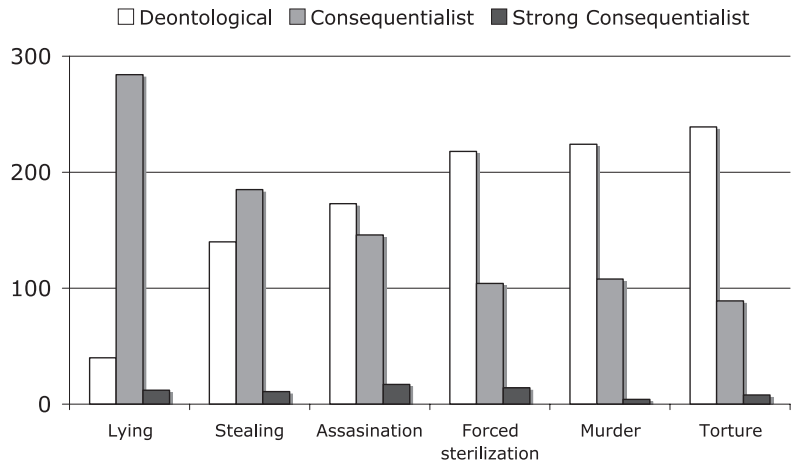

Fig. 1. Number of participants (of 336) selecting each of the three options in the *commitments* task for each queried action.

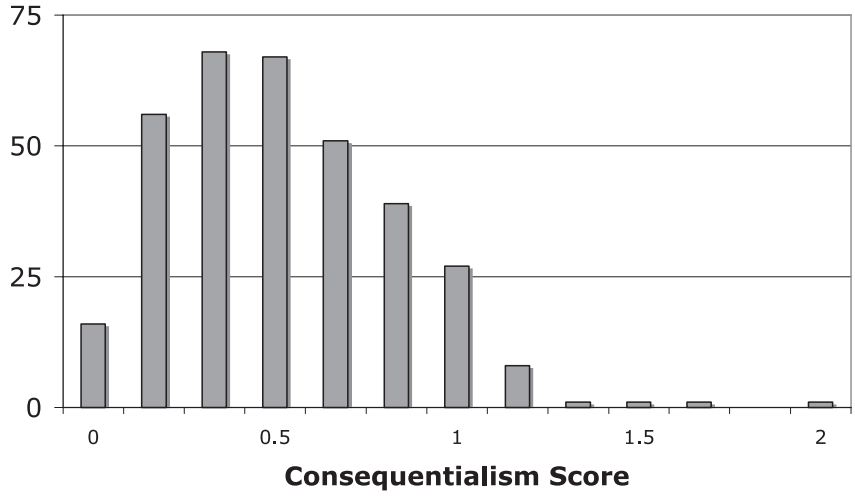

Fig. 2. Distribution of consequentialism scores. Higher values indicate more consequentialist responses.

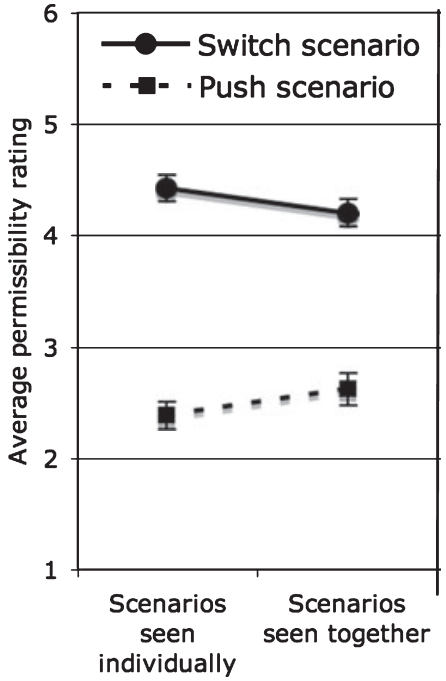

Fig. 3. Average permissibility ratings for the switch and push scenarios as a function of condition.

in the *Switch* and *Push* conditions [mean permissibility of 4.42 (1.26) vs. 2.39 (1.30), $t(222) = 11.85$, $p < .01$, independent samples t-test, $r = .62$] or within subjects in the *Both* condition [4.20 (1.31) vs. 2.63 (1.53), $t(111) = 10.27$, $p < .01$, paired-samples t-test, $r = .70$]. Responses did not vary as a function of task order in the *Switch* condition

[$t(110) = .53$, $p = .60$], the *Push* condition [$t(110) = .15$, $p = .89$], or the *Both* condition [$t(110) = .43–.56$, $p = .58–.67$]. However, permissibility judgments for the switch scenario in the *Both* condition did vary as a function of scenario order (see also Petrinovich & O'Neill, 1996), with subjects who saw the switch scenario first providing higher permissibility ratings than those who saw it after the push scenario [4.59 (1.11) vs. 3.80 (1.39); $t(110) = 3.30$, $p < .01$, $r = .30$]. Unlike responses on the *commitments* task, there were no significant sex differences in responses on the switch scenario [men: 4.30 (1.31), women: 4.31 (1.38), $t(222) = -.10$, $p = .92$] nor on the push scenario [men: 2.42 (1.49), women: 2.56 (1.38), $t(222) = -.69$, $p = .49$].

Seeing the switch and push scenarios side by side may have led participants to provide more similar permissibility ratings for the two scenarios than when the scenarios were presented individually. To examine this possibility, a single difference score was computed for participants in the *Both* condition, consisting of their permissibility rating on the switch scenario minus their permissibility rating on the push scenario. The mean value for this difference score in the *Both* condition was compared with the difference between the mean response to the switch scenario in the *Switch* condition and the mean response to the push scenario in the *Push* condition. Evaluating whether this linear combination of means differed from zero, as in a contrast analysis, revealed a significant effect [$t(316) = 2.03$, $p < .05$, equal variances not assumed, $r = .11$]. As Fig. 3 illustrates, participants in the *Both* condition tended to find the switch scenario less permissible and the push scenario more permissible than participants in the respective between-subjects conditions.

2.3. Relationship between tasks: Consequentialism score and permissibility ratings

Before examining the relationship between responses on the *commitments* task and on the *judgments* task, it is worth repeating that the order of the two tasks was counterbalanced and no order effects were found. This demonstrates that completing an initial task had no effect on the second task, and it thereby suggests that relationships between the two tasks are unlikely to result from perceived task demands or deliberate attempts at consistency.

If there is consistency across moral commitments and moral judgments, consequentialism scores should positively correlate with permissibility ratings on both scenarios. In the between-subjects conditions (*Switch* and *Push*), the Pearson correlation between consequentialism score and permissibility on the switch scenario was .22 ($p < .05$), and that on the push scenario was .17 ($p = .08$). In the within-subjects condition (*Both*), the correlations were .36 ($p < .01$) and .25 ($p < .01$), respectively. For every condition the correlation between trolley car judgments and responses on the ''murder'' item from the commitments task was numerically smaller than the correlation between trolley car judgments and the composite consequentialism score, even though the ''murder'' item is most applicable to the trolley scenarios.

A second prediction is that participants who are more consequentialist should differentiate the permissibility of the switch and push scenarios less than participants who are more deontological. For participants in the *Both* condition, there was a suggestive but unreliable negative correlation between consequentialism score and the difference between

permissibility ratings on the switch and push scenarios (Pearson's correlation, $r = -.17$, $p = .07$). To perform a comparable analysis for the between-subjects *Switch* and *Push* conditions, participants were classified as ''deontological'' ($n = 140$, 72% women) or ''consequentialist'' ($n = 196$, 63% women) depending on whether their consequentialism score was less than or at least as high as the median value of .5.[2] An ANOVA with scenario (switch vs. push) and classification (deontological vs. consequentialist) as independent variables and permissibility rating as a dependent variable did not reveal a significant interaction [$F(1, 220) = 1.21$, $p = .27$], suggesting that the difference between permissibility ratings across the two scenarios was not reliably smaller for consequentialists than for deontologists in these conditions. This analysis yielded comparable results when restricted to male participants [$F(1, 69) = .06$, $p = .80$] or to female participants [$F(1, 147) = 1.80$, $p = .18$].

2.4. Relationship between tasks: Higher-order interactions

Although consequentialists generally provided higher permissibility ratings than did deontologists on both trolley car scenarios, consequentialist participants did not generate significantly closer judgments than did deontologists in the corresponding conditions. This finding is puzzling in the *Both* condition, as one might expect the manipulation to facilitate the application of a consequentialist strategy among consequentialist participants. In the *Both* condition a larger number of consequentialists than deontologists did provide the same permissibility ratings for the switch and push scenarios (38% of subjects vs. 26%), but this difference was not significant [$\chi^2(1) = 2.06$, $p = .15$]. These predicted effects might not have been found because comparing consequentialists to deontologists in the same condition provides an inappropriate baseline. The question is not just whether consequentialists provide less discrepant ratings than deontologists who also saw both scenarios, but whether consequentialists provide less discrepant ratings in the *Both* condition than they would have had they seen the scenarios in isolation. This question can be addressed by considering a linear combination of means comparing the mean differences between ratings in the *Both* condition to the difference between the mean ratings in the *Switch* and *Push* conditions, as in the section on *Data on moral judgments*, but looking for an interaction with the additional variable of classification (deontologist vs. consequentialist). This analysis was significant [$t(268) = -2.21$, $p < .05$, equal variance not assumed, $r = .13$] and suggests that the previously reported effect of seeing both scenarios side by side versus individually was driven by consequentialists (see Fig. 4). This effect did not interact with sex [$t(118) = -27$, $p = .79$, equal variance not assumed], suggesting that it was not driven by the greater proportion of women classified as deontological. As a group, deontologists provided high permissibility ratings in the switch scenario and low permissibility ratings in the push scenario and were unaffected by whether the scenarios were presented individually or side by side. Consequentialists likewise provided high ratings for the switch scenario and low ratings for the push scenario when the scenarios were in isolation, but when the scenarios were side by side their ratings moved toward the midpoint. This effect remained marginally significant even when participants who gave the same permissibility rating on the switch and push cases were excluded [$t(293) = -1.88$, $p = .06$], which suggests that the effect was not driven by this

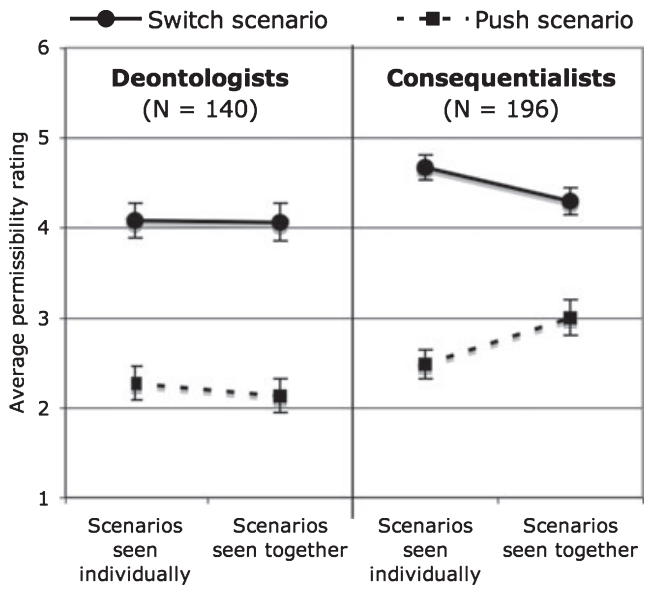

Fig. 4. Average permissibility ratings for the switch and push scenarios as a function of condition, separated by classification.

minority of participants. Rather, consequentialist participants in the *Both* condition tended to moderate their judgments for both scenarios, even if they did not ultimately provide identical permissibility ratings.

3. General discussion

Individuals vary in the extent to which their explicit moral commitments are consequentialist or deontological, and this variation is systematically related to variation in judgments elicited by trolley car problems. Specifically, consequentialist participants judge action in trolley car scenarios more permissible than do deontological participants, and only consequentialist participants moderate their judgments when the two trolley scenarios are seen side by side. In light of past findings that presenting scenarios side by side facilitates the application of rules (Gentner & Medina, 1998) and normative beliefs (Bazerman et al., 1998; O'Connor et al., 2002)—precisely the content of moral commitments—the finding that joint presentation exaggerates differences in moral judgments between consequentialist and deontological participants suggests that moral commitments partially mediate moral judgments. This pattern of findings challenges both traditional and more contemporary approaches to moral psychology.

Consistent with traditional approaches, the current findings support a relationship between moral commitments and moral judgments. However, the data challenge the idea that moral commitments are fully worked out and coherent, and that they are explicitly

applied as a basis for moral judgment. First, while participants varied in their consequentialist and deontological commitments, the distribution of commitments was not bimodal, and most participants endorsed a mix of consequentialist and deontological responses (see Fig. 2). This suggests that the consequentialist versus deontological classification reflects one or several underlying tendencies that alter the proportion of different commitments rather than tracking comprehensive frameworks of the kind moral philosophers identify with the terms ''deontology'' and ''consequentialism.'' Moreover, the extent to which the current measures of consequentialist commitments reflect reliable individual differences remains a question for future research. Second, while moral commitments were correlated with moral judgments, the relationship was at best imperfect. A majority of consequentialists distinguished the ''switch'' and ''push'' scenarios, and many judged action in these scenarios relatively impermissible. Had participants been generating moral judgments by explicitly applying moral commitments, greater correspondence would be expected.

The current findings also challenge the idea that explicit moral commitments play a negligible role in generating moral judgments. While ''dual systems'' approaches to moral psychology could in principle accommodate the current findings (or, indeed, any set of findings concerning the relationship between commitments and judgments), they would not necessarily predict them. Other work has emphasized the dissociation between trolley car judgments and the explicit reasoning reflected in justifications (Cushman et al., 2006; Hauser et al., 2007), while documenting impressive universality in trolley car judgments (Hauser, 2006; Mikhail, 2007). In particular, a large number of demographic variables, such as age, religion, and educational level, do not predict different patterns of judgment on trolley car problems (Hauser, 2006; Hauser et al., 2007), making it all the more remarkable that explicit moral commitments predict qualitative differences in the effects of joint scenario presentation. The data point toward the need for a theory of moral cognition that can account for both the associations and dissociations between explicit, abstract commitments and responses on concrete problems. This relationship will undoubtedly be complicated by unconscious principles (Cushman et al., 2006; Hauser, 2006; Mikhail, 2007), emotion (Greene et al., 2001; Haidt, 2001), and other psychological processes.

Combining insights from traditional and contemporary approaches to moral psychology, the picture that emerges is one in which moral commitments and moral judgments are systematical related, but where this relationship does not result from the explicit and consistent application of moral commitments in the course or moral judgment. In fact, a systematic relationship between moral commitments and judgments need not imply a causal role for commitments. Experience with one's own moral judgments could result in the generation of consistent commitments, or commitments and judgments could have a common cause. For example, it is possible that underlying cognitive or personality difference, such as reliance on emotion or intuition, could lead some people to more categorical moral commitments, to more deontological responses on trolley car problems, and to be unaffected by comparing scenarios (see Greene, 2007; for relevant discussions). However, this reasoning makes it likely that all participants would have generated more consequentialist judgments when the scenarios were evaluated side by side, even if the effect were attenuated among more

deontological participants (see also Greene et al., 2008). That the effect was restricted to consequntialists supports the claim that moral commitments, at least in part, mediate moral judgments.

The current findings cannot definitively rule out the possibility that moral judgments lead to moral commitments or share a common cause, but they do restore the more traditional possibility that moral commitments mediate moral judgment, albeit partially or inconsistently. For example, moral commitments could serve as a corrective: a variety of factors could influence initial judgments, with moral commitments mediating whether and how these initial judgments are adjusted. Another possibility is that moral commitments are explicitly applied in the course of judgment, but only under special circumstances, such as when actions are compared across actual or counterfactual scenarios. Evidence for differential application of theoretical commitments comes from the literature on ''sacred'' or ''protected'' values, which finds that moral commitments influence judgments about trade-offs in a task-dependent way. Those who are explicitly committed to the moral status of a good such as biodiversity tend to respond in a way that is *less* consequentialist when asked to endorse trade-offs (e.g., sacrificing one species to save five; Baron & Spranca, 1997; Tetlock, 2003), but the same participants will respond in a way that is *more* consequentialist if the judgment concerns which trade-off to make, not whether a trade-off is appropriate (Bartels & Medin, 2007). These findings can be understood if the role of moral commitments varies depending on whether a judgment concerns the choice of action or inaction in a given scenario, or different actions across scenarios.

A final possibility is that moral commitments play a causal role in judgment, but not through explicit reasoning in the course of reaching judgments (see Pizarro & Bloom, 2003; for a related suggestion). Striking evidence for this possibility comes from the finding that being a vegetarian for moral reasons—the result of explicit moral commitments—can gradually lead to feelings of disgust at the thought of eating meat (Rozin, Markwith, & Stoess, 1997). This demonstrates that moral commitments can affect more emotional and automatic responses, and that a causal relationship between commitments and judgments need not derive from the explicit and deliberate application of endorsed principles. Generalizing from vegetarianism, the suggestion is that moral commitments gradually impact the mechanisms involved in moral judgment, and thereby play an indirect causal role.

Understanding the extent and sources of plasticity in moral judgment is of particular importance, as intuitive and emotional responses presumably play a role in policing our own actions and judging those of others. This paper suggests that explicitly held moral commitments are one source of variation in moral judgment. Given that debates about moral issues occur in the language of explicit commitments, malleable commitments that impact judgment are a prerequisite to moral progress.

Notes

1. Here and in the remainder of the paper, I use the terms ''consequentialist'' and ''deontologist'' loosely. In particular, I do not consider nuanced views according to which

the act of pushing a person could have different consequences from switching a train by violating a rule, for example. I also do not intend the terms to carry the full commitments they do in philosophy, such as views about agent neutrality (see Sinnott-Armstrong, 2007, for discussion).

2. This classification achieved the most comparable group sizes possible given the distribution of consequentialism scores. Although participants are referred to as ''deontological'' or ''consequentialist,'' this should more properly be understood to mean ''less consequentialist than the median'' and ''at least as consequentialist as the median.''

Acknowledgments

The author wishes to acknowledge Fiery Cushman, Tom Griffiths, Joshua Knobe, and three anonymous reviewers for helpful comments on previous drafts, Tom Wickins for statistical advice, and Minh-Chau Do and Beibei Luo for help with data collection.

References

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.

Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices? *Psychological Science*, *18*, 24–28.

Bazerman, M. H., Moore, D. A., Tenbrunsel, A. E., Wade-Benzoni, K. A., & Blount, S. (1999). Explaining how preferences change across joint and separate evaluation. *Journal of Economic Behavior & Organization*, *39*, 41–58.

Bazerman, M. H., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of Management Review*, *23*, 225–241.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment. *Psychological Science*, *17*, 1082–1089.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5–15.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*, 263–297.

Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology*, Volume 3 (pp. 35–80). Cambridge, MA: MIT Press.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–1154.

Greene, J. D., Somerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fRMI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.

Haidt, J. (2001). The emotional dog and it rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: HarperCollins.

Hauser, M., Cushman, F., Young, L., Jin, R., & Mikhail, J. (2007). A dissociation between moral judgment and justification. *Mind & Language*, *22*, 1–21.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*, 576–590.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 151–235). New York: Academic Press.

McIntyre, A.. (2006). Doctrine of double effect. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/sum2006/entries/double-effect/.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences*, *11*, 143–152.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, *41*, 663–685.

O'Connor, K. M., De Dreu, C. K. W., Schroth, H., Barry, B., Lituchy, T. R., & Bazerman, M. H. (2002). What we want to do versus what we think we should do: An empirical investigation of intrapersonal conflict. *Journal of Behavioral Decision Making*, *15*, 403–418.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and Framing effects on moral intuitions. *Ethology and Sociobiology*, *17*, 145–171.

Petrinovich, L., O'Neill, P., & Jorgensen, M. J. (1993). An empirical study of moral intuitions: Towards an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*, 467–478.

Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychological Review*, *110*, 193–196.

Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences to values and the recruitment of disgust. *Psychological Science*, *8*, 67–73.

Singer, P. (2003). Interview: Peter Singer. *Heilpadagogik Online*, *1* (3), 49–59.

Sinnott-Armstrong, W. (2007). Consequentialism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Spring 2007 edition)*. http://plato.stanford.edu/archives/spr2007/entries/consequentialism.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*, 320–324.

Thompson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*, 1395–1415.

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

Waldmann, M. R., & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18* (3), 247–253.