

The Role of Conscious Reasoning and Intuition in Moral Judgments: Testing three principles of harm

Fiery Cushman¹, Liane Young¹ & Marc Hauser^{1,2,3}

Departments of ¹Psychology, ²Organismic and Evolutionary Biology, and ³Biological Anthropology Harvard University

ABSTRACT. *Is moral judgment accomplished by intuition or conscious reasoning? An answer to this question demands a detailed account of the moral principles in question. Here we investigate three principles guiding subjects' moral judgments and then ask whether they are invoked to explain those judgments. Across a variety of moral dilemmas, subjects' judgments about the permissibility of harming an individual aligned with three principles: (1) harm caused by action is worse than harm caused by omission, (2) harm intended as the means to a goal is worse than harm foreseen as the side-effect of a goal, and (3) harm involving physical contact with the victim is worse than harm involving no physical contact. Subjects generally appealed to the first and third principles in their justifications, but not to the second principle. This finding has significance for the methods and theories of moral psychology: the moral principles used in judgment must be directly compared to those articulated in justification and, when they are, evidence emerges that some moral principles are available to conscious reasoning while others are not.*

(in press) *Psychological Science*

INTRODUCTION

A topic of recent concern in moral psychology is the extent to which conscious reasoning, as opposed to intuition, plays a role in determining moral judgment (Greene & Haidt, 2002; Haidt, 2001; Pizarro & Bloom, 2003). In terms common to social psychology, the question is whether moral judgment is a controlled or an automatic process (Bargh, 1999).

The conscious reasoning perspective has been the central focus for students of moral development in the tradition of Kohlberg (1969). Kohlberg characterized the child's moral development by focusing on the content of her justifications rather than the source of her moral judgments. An implicit assumption of this perspective is that we generate moral judgments via the principles we articulate in moral justification, a theoretical position we refer to as "conscious reasoning".

One challenge to the conscious reasoning perspective comes from work by Haidt in which subjects fail to articulate sufficient justifications for their moral judgments (Haidt & Hersh, 2001). Haidt proposes that moral judgments arise as intuitions generated by automatic cognitive processes, and that the primary role of conscious reasoning is not to generate moral judgments, but to provide a post-hoc basis for moral justification (Haidt, 2001). We refer to the theoretical position that moral judgment is accomplished without conscious reasoning as "intuitionism". Although there is increasing support for the role of intuition in moral judgment, some argue that both "reason" and "intuition" play a role in judgment, as well as justification (Greene, *in press*); Pizarro & Bloom, 2003; Pizarro *et al.*, 2003).

A critical ingredient missing from the current debate is an experimental method that clearly links data on moral judgment with data on moral justification. Without establishing that an individual uses a specific moral principle, it makes little sense to ask whether the content of that principle is directly available to conscious reasoning. Therefore, the present study first identifies three moral principles used by subjects in the judgment of moral dilemmas, and then explores the extent to which subjects generate justifications based on these principles. Our approach, adopted in part from moral philosophy, was to compare judgments across tightly controlled pairs of scenarios. We parametrically varied each pair of scenarios to target only one factor at a time, holding all others constant. We use the term 'principle' to denote the consistent distinctions that subjects make between cases that can be attributed to a single factor, without making assumptions about the underlying psychological mechanism.

We investigated three principles:

The Action Principle: Harm caused by action is morally worse than equivalent harm caused by omission.

The Intention Principle: Harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side-effect of a goal.

The Contact Principle: Using physical contact to cause harm to another is morally worse than causing equivalent harm to another without physical contact.

The action principle is well-researched in the psychology literature, where it is often called 'omission bias' (Baron & Ritov, 2004; Spranca *et al.*, 2003). The relevance of the action principle is also recognized in the philosophical literature (Quinn, 1989; Rachels, 1975). The intention principle, often identified as the Doctrine of the Double Effect, has received intense scrutiny by philosophers (Foot, 1967; Nagel, 1986), but markedly less by psychologists (but see (Mikhail, 2002; Royzman & Baron, 2002)). The contact principle is comparatively understudied in both psychology and philosophy; although it bears a superficial similarity to Greene's 'personal-impersonal' distinction (Greene *et al.*, 2001), physical contact is neither a necessary nor a sufficient condition for a personal moral dilemma.

Having established that subjects make use of a principle, we can then ask whether this principle is available to conscious reflection during justification. We hypothesize that a hallmark of conscious reasoning is that the principles used in judgments are articulated in justifications. On the other hand, we hypothesize that intuitive responses are accompanied by insufficient justifications, uncertainty about how to justify, denial of a relevant principle, or confabulation of alternative explanations for behavior. While it is possible that moral principles consistently cited during justification were nonetheless engaged without conscious reasoning during judgment, it may be concluded that these principles are at least available for

conscious processes of moral reasoning. By contrast, those principles that consistently cannot be cited during justification would appear to be genuinely inaccessible to conscious processes of reasoning.

METHODS

Subjects voluntarily logged on to the Moral Sense Test website, moral.wjh.harvard.edu. Previous work with a different set of dilemmas revealed no substantive differences in responses obtained from subjects on the web and in more traditional, pen and paper tests (Hauser *et al.*, in press-a). Subjects averaged 37 years old with a small male bias (58%). We instructed subjects to participate only if they were fluent in English; 88% listed English as their primary language, with the United States, Canada, and the United Kingdom comprising the dominant nationality. Twenty-five percent had been exposed to formal education in moral philosophy.

After completing a demographic questionnaire, subjects received 32 moral dilemmas and 2 control cases, separated into two 'blocks' of 16. Order of presentation was counterbalanced between subjects. For each dilemma, subjects rated the protagonist's harmful action or omission on a scale from 1 to 7, with 1 labeled "Forbidden", 4 labeled "Permissible", and 7 labeled "Obligatory". In a third block, subjects were asked to justify their pattern of responses for up to five pairs of scenarios. We only asked subjects to justify responses conforming to the three principles tested (e.g. when an action was judged worse than an omission). Subjects were presented with the text of the two scenarios side-by-side, reminded which they judged more permissible, and asked to justify their pattern of responses.

All subjects had the opportunity to exit the testing session after any number of blocks. We analyzed data only from subjects who successfully completed all three blocks. Subjects were omitted from all analyses if they failed either of two control questions (by judging permissible the killing or allowed death of 5 people despite a costless alternative), or if they completed any of the 34 scenarios in fewer than 4 seconds, deemed the minimum comprehension and response time in pilot research. Additionally, subjects were removed from the analyses of justifications if they misunderstood the task, provided a nonsensical response, or provided a judgment that made it clear they had misunderstood a scenario. These subjects were not removed from our judgment analyses because not every subject justified each judgment, thereby precluding the uniform application of this procedure. Among 591 justifications, 65 were removed from the analysis of justifications.

The 32 test scenarios were arranged into 18 controlled pairs. What follows are brief descriptions of four scenarios; the actual text of all 32 scenarios is available at moral.wjh.harvard.edu/methods.html.

Evan (action, intended, no contact): Is it permissible for Evan to pull a lever that drops a man off a footbridge and in front of a train in order to cause the man to fall and be hit by the train, thereby slowing it and saving five people ahead on the tracks?

Jeff (omission, intended, no contact): Is it permissible for Jeff not to pull a lever that would prevent a man from dropping off a footbridge and in front of a train in order to allow the man to fall and be hit by the train, thereby slowing it and saving five people ahead on the tracks?

Frank (action, intended, contact): Is it permissible for Frank to push a man off a footbridge and in front of a train in order to cause the man to fall and be hit by the train, thereby slowing it and saving five people ahead on the tracks?

Dennis (action, foreseen, no contact): Is it permissible for Dennis to pull a lever that redirects a trolley onto a side track in order to save five people ahead on the main track if, as a side-effect, pulling the lever drops a man off a footbridge and in front of the train on the side tracks, where he will be hit?

Some scenarios belonged to more than one pair; for instance, "Evan" was contrasted with "Jeff" to yield an action principle comparison, with "Frank" to yield a contact principle comparison, and with "Dennis" to yield an intention principle comparison (Figure 1). Six pairs varied across the action principle, six across the intention principle, and six across the contact principle.

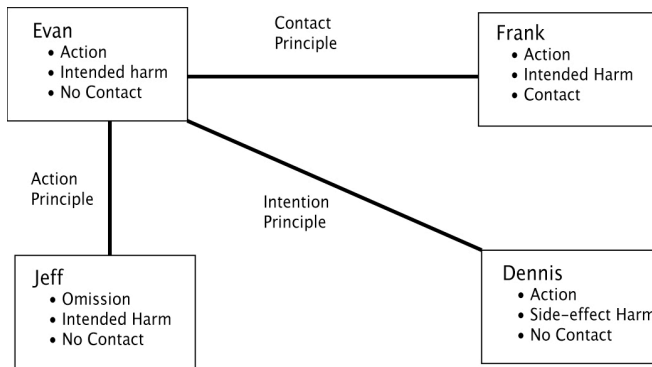


Figure 1. Principle-based contrasts between 4 scenarios arranged into 3 controlled pairs

The methods used were in accordance with the regulations of the institutional review board at Harvard University.

RESULTS

Judgments

Paired-sample t tests were performed on each of the 18 controlled pairs of scenarios to determine whether subjects rated one scenario in the pair significantly more permissible than the other in the direction predicted by the relevant principle. Statistical significance was achieved in 17 out of 18 pairs at $p < .05$, two-tailed, $N = 332$; in the remaining pair, mean permissibility ratings trended in the appropriate direction but fell short of significance, $p = .144$ (Table 1). Across scenarios with different content, subjects judged action as worse than omission, intended harm as worse than foreseen harm, and harm via contact as worse than harm without contact.

TABLE 1
Paired Samples Tests for 18 Controlled Pairs of Moral Scenarios

		Mean Difference	Std. Dev.	t	Effect Size	Sig. (2-tailed)	Probability of Replication
Action Principle Pairs	Boxcar	0.70	2.03	6.32	$d = 0.34$	$p < 0.001$	$p \text{ rep} > 0.999$
	Pond	1.69	2.00	15.34	$d = 0.84$	$p < 0.001$	$p \text{ rep} > 0.999$
	Ship	0.83	2.01	7.56	$d = 0.41$	$p < 0.001$	$p \text{ rep} > 0.999$
	Car	0.90	1.77	9.26	$d = 0.50$	$p < 0.001$	$p \text{ rep} > 0.999$
	Boat	0.98	1.98	8.97	$d = 0.49$	$p < 0.001$	$p \text{ rep} > 0.999$
	Switch	0.26	1.87	2.56	$d = 0.13$	$p = 0.011$	$p \text{ rep} = 0.942$
Intention Principle Pairs	Speedboat	0.29	1.15	4.65	$d = 0.25$	$p < 0.001$	$p \text{ rep} > 0.999$
	Burning	1.12	1.58	12.90	$d = 0.70$	$p < 0.001$	$p \text{ rep} > 0.999$
	Boxcar	0.50	1.68	5.38	$d = 0.29$	$p < 0.001$	$p \text{ rep} > 0.999$
	Switch	0.28	1.77	2.92	$d = 0.15$	$p = 0.004$	$p \text{ rep} = 0.970$
	Chemical	0.24	1.51	2.91	$d = 0.15$	$p = 0.004$	$p \text{ rep} = 0.970$
	Shark	0.30	1.77	3.14	$d = 0.16$	$p = 0.002$	$p \text{ rep} = 0.979$
Contact Principle Pairs	Speedboat	0.89	1.44	11.36	$d = 0.62$	$p < 0.001$	$p \text{ rep} > 0.999$
	Burning (1)	0.24	1.40	3.18	$d = 0.17$	$p = 0.002$	$p \text{ rep} = 0.979$
	Boxcar	1.07	1.72	11.28	$d = 0.62$	$p < 0.001$	$p \text{ rep} > 0.999$
	Burning (2)	0.37	1.22	5.50	$d = 0.30$	$p < 0.001$	$p \text{ rep} > 0.999$
	Aquarium	0.17	1.35	2.31	$d = 0.12$	$p = 0.021$	$p \text{ rep} = 0.948$
	Rubble	0.10	1.27	1.47	$d = 0.07$	$p = 0.144$	$p \text{ rep} = 0.774$

Note. Degrees of freedom were 331 for all tests. Probability of replication ($p \text{ rep}$) was calculated according to Killeen (2005).

Justifications

A total of 526 justifications were coded for five non-exclusive attributes. The attributes were:

Sufficient: The subject mentioned any factual difference between the two cases and either claimed or implied that it was the basis of his judgments. It was not necessary for the subject to identify

the target principle, so long as the principle generated by the subject could adequately account for his pattern of responses on the scenario pair in question.

Failed: The subject suggested an alternative principle, but this alternative could not account for his actual pattern of judgments.

Uncertainty: The subject explicitly referenced his own uncertainty about how to justify the scenario, or directly stated that he could not justify his response.

Denial: The subject stated that, at the point of justification, he did not consider there to be any moral difference between the two scenarios.

Alternative explanation: The subject appealed to an alternative explanation of his response, either [1] invoking facts that were not present in the scenarios or [2] claiming that he made a mistake in selecting the appropriate response, for instance, by clicking on the wrong button.

Our complete coding criteria, including examples from the data, are available at moral.wjh.harvard.edu/methods.html. Justifications were coded by FC and a colleague familiar with the research, and ties were broken by LY. The overall percent agreement between coders for individual attributes are provided in Table 2, along with Cohen's kappa, a statistic of inter-observer reliability for which values between 0.60 and 0.70 are considered fair, 0.70 to 0.75 good, and above 0.75 excellent (Fleiss, 1981). For one of the attributes, "Failed", the low Cohen's kappa of .32 warrants caution in the interpretation of results, although the overall percent agreement of 89% was quite high.

Differences between principles in the proportion of justifications meeting criteria for each attribute were tested by chi-square analysis (Table 2). For all five attributes, the proportion of justifications meeting criteria differed significantly between principles. Critically, subjects readily provided sufficient justifications for the action principle, rarely did so for the intention principle, and showed an intermediate level of responding for the contact principle. While justifications were coded as sufficient even when subjects provided logically adequate justifications for their behavior other than the target principle in question, we observed that 95% of sufficient justifications depended on the target principle.

TABLE 2
Proportions of Justifications Exhibiting Attributes, by Principle

		Proportions, by Principle			Chi-Square Analysis			Inter-Observer Reliability	
		Action	Intention	Contact	χ^2	Sig.	Cramer's ϕ	OPA	Kappa
Attribute	Sufficient	0.81	0.32	0.60	91.149	<.001	0.42	89%	0.77
	Failed	0.06	0.16	0.10	10.869	0.004	0.14	89%	0.32
	Uncertain	0.05	0.22	0.04	39.058	<.001	0.27	97%	0.78
	Denial	0.02	0.17	0.13	28.781	<.001	0.23	95%	0.67
	Alternative Explanation	0.10	0.29	0.32	34.344	<.001	0.26	90%	0.69

Note. This table displays the proportion of justifications of scenario pairs targeting a particular principle that met criteria for a particular attribute, and chi-square analyses test for significant differences between proportions. All chi-square analyses were conducted with N = 526 and 2 degrees of freedom.

It is possible that the attribute "alternative explanations" represents a baseline error rate for the task—that is, subjects' alternative explanations are actually true accounts of their behavior. If this interpretation is accurate, then the proportion of alternative explanations out of the total number of scenario-pairs should be equal across the three principles. (This analysis differs from the one presented above where we present the proportion of alternative explanations out of the subset of scenario-pairs for which subjects' judgments conformed to the predicted pattern). Analyses revealed statistically significant differences among these proportions (χ^2 (2, N = 1502) 10.163, p=.006). Subjects were almost twice as likely to arrive at an alternative explanation after judging a pair of intention principle (9%) or contact principle (9%) scenarios than after judging a pair of action principle scenarios (5%). Alternative explanations therefore seem to be driven by the principle demanded in justification, and do not exclusively represent a baseline error rate for the task.

A more stringent test of the differences in the attributes elicited by each of the three principles treats controlled pairs of scenarios as the unit of analysis, rather than individual justifications. After calculating for each scenario-pair the proportion of justifications meeting criteria for a given attribute, we used an ANOVA to determine whether the mean proportion for each attribute differed by principle. Though this reduced the sample size from 526 justifications to 18 scenario-pairs, each ANOVA confirmed the previous set of analyses (Table 3). To rule out the hypothesis that differences between principles

were driven by the size of the judgment difference between items in a scenario-pair, we repeated the ANOVAs with mean judgment differences as a covariate. Significant differences by principle remained for all five ANOVAs.

TABLE 3
Differences in Attributes by Principle, and ANOVA analyses

	Mean Proportions, by Principle			ANOVA		
	Action	Intention	Contact	F	Sig.	partial η^2
Sufficient	0.79	0.30	0.59	22.44	<.001	0.749
Failed	0.07	0.18	0.09	5.75	0.014	0.434
Uncertain	0.05	0.23	0.04	12.28	0.001	0.621
Denial	0.02	0.20	0.14	8.14	0.004	0.520
Alternative	0.11	0.29	0.34	8.60	0.003	0.534
Explanation						

Note. Degrees of freedom are (2,15) for all analyses.

DISCUSSION

This study presents evidence for three principles that guide judgments of moral dilemmas, each focused on tradeoffs of life and death, but varying in content. Subjects rated harmful actions as morally worse than harmful omissions (the action principle), harm intended as the means to an end morally worse harm foreseen as the side-effect of an end (the intention principle), and harm involving physical contact as morally worse than harm without contact (the contact principle).

The content of subjects' justifications differed greatly by principle. In the case of the action principle, a large majority of subjects were able to provide sufficient justifications for their judgments, while relatively few provided failed justifications, denied any moral difference between the scenarios, or expressly doubted their ability to justify their responses. These data are consistent with the conscious reasoning model of moral judgment. Although it is possible that subjects constructed the action principle post-hoc upon having to justify their responses, a large majority of subjects at least possess the requisite explicit knowledge required by the conscious reasoning account.

Quite the opposite is true of the intention principle. Less than a third of subjects were able to provide sufficient justifications for their pattern of judgments. In 22% of justifications subjects specifically indicated uncertainty about how to justify their responses, in 17% subjects denied that there was any morally relevant difference between the cases, and in 16% subjects provided justifications that in fact failed to account for their pattern of judgments. These data are more consistent with an intuitionist model of moral judgment. Subjects reliably generated a pattern of moral judgments consistent with the intention principle, but were generally incapable of articulating it, to the point of expressing confusion when confronted with their own judgments or even denying their judgments altogether. The intention principle clearly plays a role in moral judgment, but most likely in the form of unconscious—or at least inexpressible—knowledge.

It is notable that subjects were more than three times as likely to have evoked alternative explanations for their responses in intention cases as compared to action principle cases, either by claiming to have made an error (e.g., by pushing the wrong button) or to invoke unwarranted assumptions about the scenarios (e.g., assuming that a harm would actually not occur, even when the scenario explicitly stated that it would). There is no reason to believe that the particular scenarios used in intention principle pairs should be any more likely to invite either errors or assumptions than those used in action or contact principle pairs—the scenarios were identically constructed, and indeed in several cases individual scenarios were used in contrasts of both types. Thus, what determines the use of alternative explanations appears not to be the scenario presented during judgment, but rather the principle targeted in justification. We conclude that in some cases subjects' appeals to alternative explanations are confabulations generated at the moment of justification and prompted by the subject's inability to justify their pattern of judgments. Parallel evidence comes from Wheatley and Haidt's (2005) recent priming study under hypnosis where subjects' confabulation accompanies their inability to provide a principled justification of moral judgment.

Subjects' justifications of their responses to contact principle cases occupy an intermediate position between the action principle and the intention principle. Subjects were typically able to articulate the relevant principle used, but unwilling to endorse it as morally valid. Sixty percent of justifications in contact principle cases were sufficient, nearly twice the proportion for intention principle cases. Additionally, only 10% of justifications were classified as failed, and 4% indicated uncertainty. These statistics are all comparable to those corresponding to action principle cases. Yet, in 13% of contact principle justifications, subjects denied that physical contact made a moral difference, a proportion much closer to the intention principle cases (17%) than the action principle cases (2%). Subjects were also about three times as likely to have appealed to alternative explanations in cases of the contact principle than in cases of the action principle—a nearly identical pattern of results to the intention principle. Thus, although subjects are able to articulate the principle behind their reasoning, they often reject it as morally invalid, or confabulate alternative explanations for their behavior. For instance, one subject wrote, "I guess I was fooled by the line-pulling seeming more passive than the man-pushing, but that view is hard to justify now".

Although a conscious reasoning interpretation of subjects' justifications for contact principle cases cannot be definitively rejected, the data favor the intuitionist view. The observation that many subjects used but were unwilling to endorse the contact principle is not readily explained by the conscious reasoning model. Why would a subject reason consciously from an explicit principle about physical contact during judgment, but then disavow the same principle during justification? A more plausible explanation is that the contact principle guides moral judgments according to the intuitionist model during judgment, and that a process of post-hoc reasoning at justification allows subjects to deduce the principle behind their judgments. Once deduced, the principle is regarded as morally irrelevant. Just such a process of post-hoc reasoning has been proposed by Haidt (2001). This raises the question of why a similar process is not observed for the intention principle cases. One possibility is that the features picked out by the intention principle are less salient than those picked out by the contact principle; this speculation requires additional research.

In summary, our results show that while some moral principles are available in a large majority of subjects for conscious reasoning, others are not available and appear to operate in intuitive processes. The extent to which one of these processes plays a more dominant role depends on the particular moral principles triggered: the intention principle seems best characterized by the intuitionist model, the contact principle by intuitive judgment followed by rational reflection, and the action principle either by conscious reasoning or, at a minimum, post-hoc conscious reasoning. These data bear in important ways on questions central to the study of moral cognition. We briefly raise three such issues.

First, our results underscore the methodological importance of linking subjects' expressed moral principles to the operative principles underlying their patterns of judgment (Hauser, in press; Mikhail, 2000). Controlled pairs of scenarios have long been used in philosophy and psychology to probe moral judgments (Kamm, 2001; Petrinovich *et al.*, 1993; Spranca *et al.*, 1991; Thompson, 1985), but have not been extended to moral justification. Without using tightly controlled scenario pairs to target narrowly defined principles, it is not possible to determine whether a subject's justification can account for her judgments. It has been demonstrated that under uniform testing conditions this method can yield data that distinguishes principles available to conscious reasoning from those that are not. By demonstrating divergent results under a common paradigm, we are better situated to detect genuine differences in the application of moral principles.

This methodological point has been underappreciated by advocates of both conscious reasoning and intuitionist models. Within the tradition established by Kohlberg, the moral principles cited in justification were simply assumed to be identical to the moral principles used in judgment. The high rate of failed justifications for the intention principle cases reported here suggests that this assumption is not always valid. On the other hand, Haidt (2001) has demonstrated moral dumbfounding by asking subjects why, for instance, it is wrong for a brother and sister to have intercourse without demonstrating that subjects' patterns of judgments necessitate a principle more complex than "it is wrong for a brother and a sister to have intercourse". In our own work, dumbfounding might arise if we were to ask subjects *why* a harmful action is worse than a harmful omission. But this is an altogether different question from whether subjects have explicit knowledge of the principle to which their pattern of judgments conforms, which is simply that a harmful act *is* less permissible than a harmful omission. From either a conscious reasoning or an intuitionist perspective, a proper analysis of subjects' justifications cannot be conducted in the absence of a precise accounting of their judgments. However, because of the possibility of post-hoc

reasoning in cases where subjects provide sufficient justifications, the present method is better suited to rejecting the role of conscious reasoning than supporting it. Future research must design methodologies that provide strong evidence in favor of consciously reasoned moral judgments.

Second, our results provide evidence for moral principles that are formulated over detailed representations of the causal and intentional aspects of harm. To date, many of the moral principles proposed by psychologists have far broader content than those we present. Kohlberg's moral stages are formulated over concepts like authority, cooperation, and autonomy (Kohlberg, 1969). Greene proposes an emotion-based appraisal system with a general prohibition of actions fitting the semantic structure "me hurt you" (Greene *et al.*, 2004). Haidt formulates moral principles over the concepts of suffering, reciprocity, hierarchy, purity and group identity (Haidt, in press). Ultimately, these useful generalizations will have to be married to a more detailed analysis of the basic concepts such as action, intention, and contact (Hauser *et al.*, in press-a; Hauser *et al.*, in press-b; Knobe, 2003; Mikhail *et al.*, 2002; Pizarro *et al.*, 2003).

Finally, our results support the view that moral judgment can be accomplished by multiple systems: some moral principles are available to conscious reflection—permitting but not guaranteeing a role for conscious reasoning—while others are better characterized by an intuitionist model. On the basis of neuroimaging data, Greene and colleagues (2004) have proposed a two-system model for moral judgment, one characterized by the engagement of affective systems and the other by the engagement of cognitive systems. These data suggest that, regardless of where the division between affect and cognition is placed, a multi-system model of moral judgment is warranted. Such a multi-system model stands in contrast to Kohlberg's perspective, in which all moral reasoning is assumed to be the product of conscious reasoning. It also stands in contrast to the recent intuitionist proposal by Mikhail to incorporate the sorts of moral principles discussed here into a single evaluative mechanism analogous to the perception of syntactic structure in language (Mikhail *et al.*, 2002).

In conclusion, this paper provides a novel methodological approach to the study of our moral psychology, highlighting the interaction of intuition and conscious reasoning and emphasizing the distinction between the principles that people use and the principles that people articulate.

ACKNOWLEDGEMENTS

We thank Jennifer Pegg for her contribution to the analysis of justification data, and Jonathan Baron, Jonathan Haidt, David Pizarro and Joshua Knobe for their comments on an earlier version of this manuscript.

REFERENCES

- Bargh, J. A. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462-479.
- Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, 94, 74-85.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd Edition ed.). New York: Wiley.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- Greene, J. D. ((in press)). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology and biology*. New York: Oxford University Press.
- Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, 6, 517-523.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389-400.
- 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. (in press). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology*. Oxford: Oxford University Press.
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31(1), 191-221.
- Hauser, M. D. (in press). *Moral minds: How nature designed a universal sense right and wrong*. New York: Harper Collins.
- Hauser, M. D., Cushman, F. A., & Young, L. (in press-a). A dissociation between moral judgment and justification. *Mind and Language*.
- Hauser, M. D., Young, L., & Cushman, F. A. (in press-b). Reviving Rawls' linguistic analogy. In W. Sinnott-Armstrong (Ed.), *Moral psychology and biology*. New York: Oxford University Press.
- Kamm, F. M. (2001). *Morality, mortality: Rights, duties, and status*. New York: Oxford University Press.
- Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, 16(5), 345-353.
- Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, 16, 309-324.
- 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.
- Mikhail, J. (2002). Aspects of a theory of moral cognition: Georgetown University Law Center.
- Mikhail, J. M. (2000). *Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'a theory of justice'*. Unpublished PhD, Cornell University, Ithaca.
- Mikhail, J. M., Sorrentino, C., & Spelke, E. (2002). *Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery, the rescue principle, the first principle of practical reason, and the principle of double effect*. Unpublished manuscript, Stanford.
- Nagel, T. (1986). Ethics. In T. Nagel (Ed.), *The view from nowhere*. Oxford: Oxford University Press.
- 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(3), 467-478.
- Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychol Rev*, 110(1), 193-196; discussion 197-198.
- Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39(653-660).
- Quinn, W. S. (1989). Actions, intentions, and consequences: The doctrine of doing and allowing. *The Philosophical Review*, 145.
- Rachels, J. (1975). Active and passive euthanasia. *New England Journal of Medicine*, 292, 78-80.
- Rozman, E., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15(2), 165-184.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27(1), 76-105.

- Spranca, M., Minsk, E., & Baron, J. (2003). Omission and commission in judgment and choice (pp. 31): University of Pennsylvania.
- Thompson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 279.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16(10), 780-784.