

## Moral Judgment

Jennifer Nado  
Daniel Kelly  
Stephen Stich

Rutgers University

### 1. Introduction

Questions regarding the nature of moral judgment loom large in moral philosophy. Perhaps the most basic of these questions asks how, exactly, moral judgments and moral rules are to be defined; what features distinguish them from other sorts of rules and judgments? A related question concerns the extent to which emotion and reason guide moral judgment. Are moral judgments made mainly on the basis of reason, or are they primarily the products of emotion? As an example of the former view, Kant held all moral requirements to be derived from a principle of rationality (the categorical imperative). As an example of the latter, Hume famously claimed that reason is “the slave of the passions” and that moral judgments stem from the moral emotions.

When addressing these issues, philosophers have largely relied on the traditional tools of philosophical analysis, along with introspection, anecdotal evidence and armchair speculation. In recent years, however, a rich body experimental psychology has emerged which, in the view of a growing number of philosophers, casts important new light on these venerable questions. Our aim, in this chapter, is to illustrate how empirical methods can help move traditional philosophical debates forward in interesting and important ways. Since space does not permit an exhaustive survey of the relevant experimental work, we will focus on a few of the most compelling examples.<sup>1</sup>

### 2. The Definition of Morality

In 1957, Alasdair MacIntyre wrote: “The central task to which contemporary moral philosophers have addressed themselves is that of listing the distinctive characteristics of moral utterances” (MacIntyre 1957). Thirteen years later, MacIntyre’s article was reprinted in an anthology called *The Definition of Morality* (Wallace & Walker 1970) which also included papers by such leading figures as Elizabeth Anscombe, Kurt Baier, Philippa Foot, William Frankena and Peter Strawson. All of these, in one way or another, tackled the question of how ‘morality’ is best defined. As one might expect from this distinguished list of authors, many of the arguments developed in the book are subtle and sophisticated. And as one might expect in just about any group of 13 philosophers, no consensus was reached. In addition to debate

---

<sup>1</sup> For more extensive reviews of the literature, see Doris & Stich (2005, 2006).

about how the notions of moral utterance, moral rule and moral norm are to be defined, many of the contributors to the volume also discuss a cluster of meta-philosophical questions, such as “What is a definition of morality supposed to do?” and “What counts as getting the definition right?” Here again, no consensus was reached.

A few years later, Paul Taylor published a long paper whose goal was to elaborate and defend an account of what it is for a norm to be a *moral* norm for a group of people (Taylor 1978). Taylor also provides a useful taxonomy of various positions one might take on the meta-philosophical issue. What, Taylor asks, might philosophers be trying to do when they offer a definition of ‘morality’ or ‘moral rule’? One option is that they are offering a *linguistic analysis* which tries to capture how the word ‘moral,’ or phrases like ‘moral rule’ and ‘moral norm’ are used by English speakers. A second, closely related possibility is that they are proposing a *conceptual analysis*, aimed at making explicit the concept of morality held by people in our society. A third, quite different alternative is that philosophers are trying to specify the *essence* of morality. Philosophers pursuing this project would maintain that moral rules or norms constitute a *natural kind* whose members all share some essential property or set of properties. The goal of the project is to discover what those essential properties are. Taylor suggests that this is a misguided project, since he believes that “there is no such essence” (52), though he recognizes that others might disagree with this assessment.<sup>2</sup>

At about the same time, a group of developmental psychologists who had been influenced by some of the philosophical literature aimed at defining morality began developing and defending their own definition. On one interpretation of their work, these psychologists were demonstrating that Taylor was wrong: morality *is* a natural kind, and via their experiments they were beginning to discover what the essential properties of moral rules are. For about two decades, this work was all but unknown to philosophers. But as the 20<sup>th</sup> century drew to a close, interest in the empirical study of morality increased dramatically among philosophers, and this work became increasingly influential.

The central figure in this research tradition is Elliot Turiel, who proposed a definition of ‘moral rule’ together with a definition of ‘conventional rule’ – another notion on which philosophers like David Lewis had recently lavished a fair amount of attention. Turiel did not defend his definitions using abstract philosophical arguments, however, nor did he make claims about how the words ‘moral’ and ‘conventional’ are used. Rather, he used his definitions to design psychological experiments, and those experiments produced some very extraordinary findings on moral judgment (Turiel, 1979, 1983).

The core ideas in the definitions that Turiel & his followers have offered are as follows:

---

<sup>2</sup> Philosophical discussion of the definition of morality has continued into the new millennium. See, for example, Gert (2005).

- i. Moral rules are held to have an objective, prescriptive force; they are not dependent on the authority of any individual or institution.
- ii. Moral rules are taken to hold generally, not just locally; they not only proscribe behavior here and now, but also in other countries and at other times in history.
- iii. Violations of moral rules involve a victim who has been harmed, whose rights have been violated, or who has been subject to an injustice.
- iv. Violations of moral rules are typically more serious than violations of conventional rules.

Conventional rules, on Turiel's account, have just the opposite cluster of properties. They do not have objective, prescriptive force; rather they are viewed as arbitrary or situation-dependent, and can be suspended or changed by an appropriate authoritative individual or institution. Conventional rules are often geographically & temporally local; those applicable in one community often will not apply in other communities or at other times in history. Violations of conventional rules do not involve a victim who has been harmed, whose rights have been violated, or who has been subject to an injustice, and these violations are typically less serious than violations of moral rule.

Guided by these definitions, Turiel and his associates developed an experimental paradigm that has become known as the *moral / conventional task*. In this task, participants are presented with examples of transgressions of prototypical moral rules and prototypical conventional rules, and are asked series of probe questions. Some of the questions are designed to determine whether the participants consider the action to be wrong, and if so, how serious it is. Other questions explore whether participants think that the wrongness of the transgression is "authority dependent." For example, a participant who has said that a specific rule-violating act is wrong might be asked: "What if the teacher said there is no rule in this school about [that sort of rule violating act], would it be right to do it then?" A third group of questions aim to determine whether participants think the rule is general in scope. Is it applicable to everyone, everywhere, or just to a limited range of people, in a restricted set of circumstances? Finally, participants are asked how they would justify the rule – do they invoke harm, justice or rights, or do they invoke other factors?

Early results suggested that the categories of moral and conventional rules, as defined by Turiel, are robustly psychologically significant. In experiments in which they were asked about both prototypical moral transgressions and prototypical conventional transgressions, participants' responses to the two sorts of transgression differed *systematically*, and in just the way suggested by Turiel's characterization of the distinction. Transgressions of prototypical moral rules almost always involved a victim who was clearly harmed; common examples included one child hitting another, or one child pushing another child off a swing. As Turiel's account predicted, these were judged to be more serious than transgressions of prototypical conventional rules, the wrongness of the transgression was judged not to be authority dependent, the violated rule was

judged to be general in scope, and judgments were justified by appeal to the harm they caused. By contrast, transgressions of prototypical conventional rules, such as a child talking in class when she has not been called on by the teacher, or a boy wearing a dress to school, were judged to be less serious, the rules were judged to be authority dependent and not general in scope, and judgments were not justified by appeal to harm. In the three decades after the moral / conventional task was first introduced, this pattern of results has been found in an impressively diverse range of participants differing in religion, nationality, culture and age – from 3½ years to adulthood.

What conclusions can be drawn from these results? It is not entirely clear how Turiel and his associates would answer this question, since much of their own discussion is couched in the philosophically tendentious and less than perspicuous terminology that grows out of the Piagetian tradition. Rather than getting bogged down in textual exegesis, we'll set out some conclusions that are plausible to draw from these findings, conclusions which many philosophers impressed by the results appear to accept. Since we'll have occasion to refer back to these conclusions, it will be useful to give them numbers.

- (1) Participants in moral / conventional task experiments will typically display one of the two *signature response patterns* described in the previous paragraph. Moreover, these response patterns are *nomological clusters* – there is a strong (“lawlike”) tendency for the members of the cluster to occur together.
- (2) Each of these signature response patterns is associated with a certain type of transgression. (2a) The “moral” signature response pattern is evoked by transgressions involving harm, justice, welfare or rights; (2b) the “conventional” signature response pattern is evoked by transgressions that *do not* involve harm, justice, welfare or rights.
- (3) The regularities described in (1) and (2) are pan-cultural, and they emerge quite early in development.

Since nomological clusters like those noted in (1), and generalizations like those noted in (2a) and (2b), are central in philosophical accounts of natural kinds (Boyd 2002), it is plausible to draw the further conclusion that both moral rules and conventional rules are indeed *natural kinds*, and that the essential features of the kinds are just those specified in Turiel's definitions.

It is not surprising that as the results of experiments using the moral / conventional task became more widely known, this work began to make an impact on naturalistically inclined philosophers interested in moral psychology. For, if true, the conclusion that both moral rules and conventional rules are natural kinds is profoundly important. There are, however, a growing number of empirically informed skeptics who doubt that conclusion is warranted. For the most part, the skeptics have focused on evidence that challenges (2b). There are many societies, the skeptics maintain, in which transgressions that *do not* involve harm, justice, welfare and rights fail to evoke the full

“conventional” response pattern. Rather, these transgressions evoke one or more of the signature “moral” responses. Perhaps the best known study illustrating this phenomenon was conducted by Jonathan Haidt and colleagues, who explored people’s judgments about a variety of transgressions including the following memorable example:

A man goes to the supermarket once a week and buys a dead chicken. But before cooking the chicken, he has sexual intercourse with it. Then he cooks it and eats it (Haidt, Koller & Dias, 1993, p. 617).

Haidt and colleagues found that low socio-economic status participants in both Brazil and the U.S.A. judged this, and a variety of other disgusting, but not harmful, transgressions to be authority *independent* and generally applicable, both of which are features of the signature moral response pattern. In a more recent study, Nichols (2002, 2004) explored participants’ reactions to violations of etiquette norms, comparing rules that prohibit disgusting behavior - for example, a dinner guest snorting into his water glass then taking a drink - with rules that prohibit behavior that is not disgusting, like a dinner guest drinking tomato soup out of a bowl. Nichols found that in children the disgusting etiquette transgressions evoked three of the signature “moral” responses – they were more serious than the conventional transgression, not authority dependent and generalizable to other groups – while the non-disgusting transgressions evoked all of the signature “conventional” responses. With adults, the disgusting transgressions evoked two of the signature “moral” responses – they were more serious and not authority dependent, but they did *not* generalize to other groups. In addition to challenging (2b), Nichols’ results pose a particularly clear challenge to the claim that the signature response patterns are nomological clusters, since he finds that for etiquette rules the signature “conventional” response pattern comes apart in three different ways!

These are not the only studies that raise problems for (2b).<sup>3</sup> However, we suspect that the published studies challenging (2b) may only be the tip of the iceberg. For a variety of reasons, researchers using the moral / conventional task have focused on a very narrow range of “conventional” transgressions, restricted almost entirely to those that would be readily understood by children. As a wider range of “conventional” transgressions is explored, we expect there will be many more cases that fail to support (2b). The emphasis on “schoolyard” transgressions in the literature is even more pronounced in the case of “moral” transgressions. Indeed, as this is being written (in 2006), there is only one study that explores a substantial range of harmful moral transgressions of the sort that would not be familiar to young children. In this study, Kelly et al. (2007) used probes like the following:

(A) Three hundred years ago, whipping was a common practice in most navies and on cargo ships. There were no laws against it, and almost everyone thought that whipping was an appropriate way to discipline sailors who disobeyed orders or were drunk on duty.

---

<sup>3</sup> For a more detailed survey, see Kelly & Stich (forthcoming).

Mr. Williams was an officer on a cargo ship 300 years ago. One night, while at sea, he found a sailor drunk at a time when the sailor should have been on watch. After the sailor sobered up, Williams punished the sailor by giving him 5 lashes with a whip.

Is it OK for Mr. Williams to whip the sailor?

YES

NO

On a scale from 0 to 9, how would you rate Mr. Williams' behavior?

(B) Mr. Adams is an officer on a large modern American cargo ship in 2004. One night, while at sea, he finds a sailor drunk at a time when the sailor should have been monitoring the radar screen. After the sailor sobers up, Adams punishes the sailor by giving him 5 lashes with a whip.

Is it OK for Mr. Adams to whip the sailor?

YES

NO

On a scale from 0 to 9, how would you rate Mr. Adams' behavior?

The results were quite dramatic. In this case, 52% of participants said it was OK to whip the sailor 300 years ago, but only 6% said it was OK to do it in 2004. This suggests that the participants in this study did *not* judge the rule prohibiting whipping in the contemporary world to be applicable at other times in history. The results on other probes used by Kelly et al. were similar. This is just one study, of course, and more work is needed before we can draw any definitive conclusion about whether Turiel and his followers have discovered the essential properties of moral and conventional rules. But we think it is already abundantly clear that experimental work of the sort we have recounted in this section provides a new and important tool for exploring a central question in moral philosophy.

### 3. Reason and Emotion in Moral Judgment

As noted in the introduction, another venerable issue in philosophical debates about moral judgment centers on whether emotion or reason plays the more important role. In the psychological literature, where the theories of Jean Piaget and of Lawrence Kohlberg enjoyed wide influence, the prevailing view for much of the 20<sup>th</sup> century favored reason. Though they differ over many important details, both Piaget and Kohlberg claim that children progress through different stages in their ability to make moral judgments, with each stage employing more sophisticated and complex moral reasoning than the last. Recently, however, a number of psychologists have offered accounts of moral judgment in which emotion has the upper hand.

One of the most radical and provocative of these accounts was proposed by Jonathan Haidt. According to Haidt’s “social intuitionist” model, emotional capacities involving affect and intuition do almost all of the work in generating moral judgments (Haidt, 2001). Reason, on the other hand, is relegated to the role of a lawyer or public relations agent, whose job it is to offer public, post-hoc justifications for judgments after they have been made. Figure 1 is a simplified depiction of the model Haidt defends.

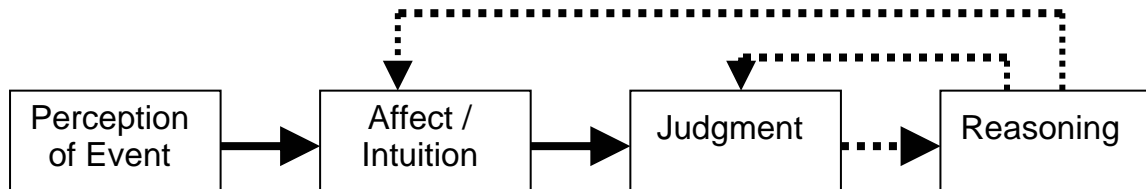


Figure 1: Haidt’s Social Intuitionist Model

The first step in the process leading to moral judgment, in this model, is the perception of a morally relevant event. The second box represents “moral intuitions,” which rapidly and spontaneously appear in consciousness in response to the witnessed moral situation. The person experiencing these intuitions normally lacks any awareness of having gone through a process of reasoning to arrive at them. Rather, Haidt characterizes these intuitions, which he holds to be the fundamental determinants of moral judgment, as affective reactions – quick flashes of disgust or anger, for instance. Often, the entire process stops once the intuition gives rise to a judgment. However, when circumstances require the person to justify her judgment, she will engage in conscious reasoning in order to produce a justification. This *post-hoc* reasoning process usually supports the affective intuition, but will occasionally override the initial affective judgment - and it may even occasionally affect the system responsible for affective intuitions. Since neither reasoning nor the downstream effects of reasoning need always occur, we’ve represented them with dashed arrows in Figure 1.

In support of this model, Haidt offers an extensive array of empirical findings. Among the most striking of these is a study in which participants were presented with vignettes, like the one that follows, which engender substantial affect but which are carefully designed to rule out most of the justifications that participants are likely to come up with.

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least, it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making

love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it okay for them to make love? (Haidt, 2001, 814)

Haidt found that participants typically answer “immediately,” insisting that the behavior was wrong. When asked why, they begin “searching for reasons” (814). But the most obvious reasons to oppose incest, like the risk of pregnancy, the higher probability of having a child with birth defects, or acquiring an unsavory reputation, do not apply in this case. When the experimenter, playing the devil’s advocate, points this out, the typical participant will readily acknowledge the point, but will still not withdraw his initial judgment. Rather, he will insist that his judgment is correct even though he cannot offer any reasons in support of that judgment. The conclusion that Haidt draws from this phenomenon, which he calls “moral dumbfounding,” is that reasoning typically plays no role in the production of moral judgment.

In another important experiment, Wheatley & Haidt (2005) hypnotized participants and told them to feel disgust when they encountered the emotionally neutral words ‘take’ or ‘often’. Participants were then asked to judge vignettes in which people behaved in morally problematic ways or in entirely unproblematic ways. Half of the participants were given versions of the vignettes with the hypnotic cue word included, while the other half received nearly identical versions of the vignettes with the hypnotic cue word omitted. This is one of the morally problematic vignettes:

Congressman Arnold Paxton frequently gives speeches condemning corruption and arguing for campaign finance reform. But he is just trying to cover up the fact that he himself [will take bribes from / is often bribed by] the tobacco lobby, and other special interests, to promote their legislation. (781)

And this is the morally neutral one:

Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion. (782)

The presence of the hypnotic cue word in the morally problematic scenarios led the participants to assess the transgressions significantly more harshly, while in the unproblematic scenarios, the presence of the cue word led a significant number of participants to judge that the agent’s actions were morally questionable! Participants were asked for comments at the end of the study and, Wheatley and Haidt report, “the post hoc nature of moral reasoning was most dramatically illustrated by the Student Council story. Rather than overrule their feelings about Dan, some participants launched an even more desperate search for external justification. One participant wrote: ‘It just seems like he’s up to something.’ ” (783)



Another account of moral judgment in which emotion plays a major role has been proposed by Joshua Greene. However, on Greene’s account, reasoning also plays a role in the production of moral judgment in an important class of cases. Greene et al. (2001) administered fMRI scans to participants while they made judgments about how people should behave when confronting a number of moral dilemmas. The dilemmas were divided into two groups. The first group involved “impersonal” moral situations like the classic “trolley problem,” where one must choose whether to flip a switch to divert a runaway trolley from a track on which it will run over five individuals to a track on which it will only kill one. The second group of dilemmas, the “personal” moral situations, included cases like the “footbridge problem” – a variation on the trolley problem where, rather than flipping a switch, one must decide whether to push an overweight man off a footbridge to stop a trolley that will kill five people if it is not stopped. The fMRI scans revealed that brain areas associated with emotion were much more active during contemplation of the personal moral dilemmas. In addition, most people judged the actions described in the personal moral dilemmas to be less permissible, and those who did judge them to be permissible took longer to make their judgments. Greene et al. believe this last finding to be a type of interference effect, where participants must suppress their tendency to judge the action impermissible.

Though Greene does not offer an explicit psychological model, his interpretation of these data suggests a model that would look something like Figure 2.

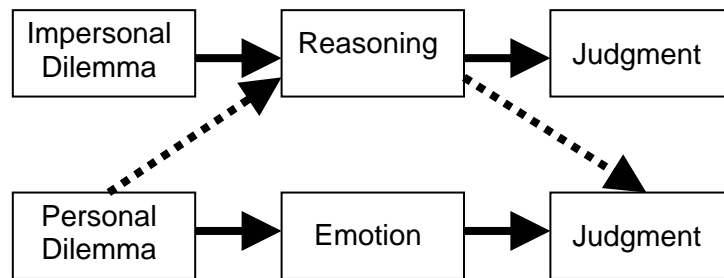


Figure 2: Greene’s Model of the Processes Underlying Moral Judgment

In this model, personal moral dilemmas trigger emotion systems, which then play a major causal role in producing a moral judgment. Impersonal moral dilemmas, however, leave the judgment to reasoning systems. The role of reasoning in personal dilemmas is either diminished or entirely absent – the dotted lines in Figure 2 represent the claim that reasoning can play a minor role in personal moral dilemmas. Although Greene’s model accords reasoning a more substantial role than Haidt’s, a central feature of both models is the heavy emphasis on the causal efficacy of emotion in the production of moral judgments.

Despite the findings of Greene and Haidt, many reject the idea that reasoning processes should be given second billing. Marc Hauser has recently argued that emotional response cannot be the primary means by which we produce our moral judgments. His own proposal is that we possess an innate, tacit capacity for moral judgment that is in many ways parallel to our capacity for language.

Hauser argues that humans are endowed with an innate ‘moral grammar’, akin to the linguistic Universal Grammar posited by Chomsky and his followers (Hauser, 2006). As in language development, this innate moral grammar provides information regarding core principles common to all moral systems. That information enables children to use cues from their environment to extract and internalize the specific moral rules present in whatever culture they are born into, even in the face of impoverished stimuli. In addition, like the linguistic faculty, the innate moral faculty operates unconsciously, quickly, and automatically. Thus, as in Haidt’s account, moral judgment is primarily intuition-based. However, Hauser denies that these intuitions are *affective*.

Hauser’s view is inspired by a passage in *A Theory of Justice* in which John Rawls suggests the use of a linguistic analogy for morality.<sup>4</sup> Hauser proposes that humans are “Rawlsian creatures” who produce moral judgments in the following manner. First, the perception of a morally significant event triggers an analysis of the actions involved. That analysis, though fast and unconscious, is a complex cognitive process in which many factors must be considered. In an important sense, it is a *reasoning* process – albeit not a conscious one. The analysis, in turn, is used to form a permissibility judgment. Emotions are triggered only after this judgment has occurred, and are relevant mainly for controlling our behavioral response to the perceived act. As in Haidt’s model, *conscious* reasoning may also come in after the initial intuitive judgment. The diagram below lays out the central features of Hauser’s view.

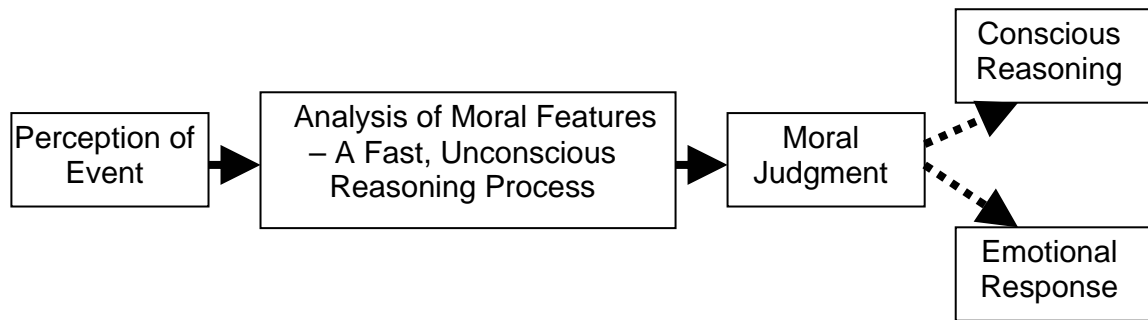


Figure 3: Hauser’s “Rawlsian” Model of Moral Judgment

Hauser contrasts the Rawlsian position with the position of those, like Haidt, who portray humans as “Humean creatures” whose emotions play a causal role in the production of moral judgments and whose reasoning capacity comes in only after the fact, and also with

---

<sup>4</sup> Rawls (1971), p 64.

those, like Piaget and Kohlberg, who suggest that humans are “Kantian creatures” whose moral judgments are largely or entirely subserved by conscious reasoning.

In arguing against the Humean creature view, Hauser notes that “neither we nor any other feeling creature can just *have* an emotion. Something in the brain must recognize – quickly or slowly – that this is an emotion-worthy situation” (Hauser 2006, p. 8). Before emotions can play any role, Hauser argues, a complex analysis of the relevant event must occur that scrutinizes the consequences, intentions, and participants involved. We must determine who did what, and why. Only then will we be equipped to make the remarkably fine-grained moral discriminations that we make.

One piece of evidence that Hauser invokes in support of his view is the fact that very slight alterations to a given situation can result in a sharp shift in permissibility judgments. Importantly, it is often difficult to account for such shifts by appeal to differences in emotional response. Hauser cites a set of trolley-problem cases developed by John Mikhail that demonstrate this phenomenon (Mikhail 2000, Mikhail et al. 2002). Recall that Greene found the emotion centers of the brain are activated during contemplation of the footbridge variant of the trolley problem. Greene hypothesized that this increased emotional reaction is responsible for our judgment that pushing the man is impermissible. Mikhail and Hauser, on the other hand, hypothesize that our innate moral grammar encodes a rule to the effect that using someone as a means to an end is wrong. Thus, pushing the man off the footbridge is impermissible because it wrongly uses the man as a means, while flipping the switch in the standard trolley case is permissible because the death of the person on the other track is a mere side-effect of the intended act of saving the five.

Mikhail presented participants with two ingenious variations on the footbridge case intended to test this hypothesis. In one case, Ned has the option of flipping a switch to divert a trolley from a track with five hikers to a looping side track containing one overweight man. If the overweight man were not present, the trolley would loop back around to the initial track and kill the five, but the overweight man is heavy enough to stop the trolley before this occurs. Thus, the overweight man is a means to saving the five. In the second case, Oscar faces a situation which is identical, except that instead of an overweight man the looping side track contains a heavy weight and a single slim hiker. The hiker is not heavy enough to stop the trolley, but the weight is; the hiker is simply in the wrong place at the wrong time, and his death will be a side effect of (rather than a means to) saving the five.

If the footbridge case is impermissible not because of its “personal” nature but because someone is used as a means, participants should judge Ned’s flipping the switch to hit the overweight man impermissible as well. This is in fact what Mikhail’s study found. In a separate study, Hauser found that, while about 90% of his participants considered deflecting the trolley in the standard trolley case to be permissible, only 50% found it permissible in Ned’s case (Hauser 2006, p. 128). This poses a problem for purely emotion-driven accounts; why should flipping a switch to hit an overweight man

on a looped track be more emotion-triggering than flipping a switch to hit a thin man on a looped track?

Though Hauser found an impressive difference between these two cases, the response on the Ned case is still quite a far cry from the response on the standard footbridge case, where only 10% deemed pushing the man off the footbridge to be permissible. However, this does not necessarily undermine Hauser's hypothesis that emotional activation plays no causal role in permissibility judgments. Hauser notes that up-close and personal moral dilemmas may trigger our moral faculty differently than action-at-a-distance cases. Consider two cases due to Peter Unger (1996). Most people judge it impermissible to leave a bleeding man lying on the side of the road even though taking him to the hospital would cause \$200 worth of damage to your car's upholstery. However, few people consider it obligatory to donate \$50 to UNICEF, even if doing so would save 25 children's lives. Hauser argues that we sense a moral difference in these two cases because "in our [evolutionary] past, we were only presented with opportunities to help those in our immediate path.... The psychology of altruism evolved to handle nearby opportunities, within arm's reach" (Hauser 2006, p. 10). A similar explanation may apply to the difference between Ned's case and the footbridge dilemma.

We are inclined to think that Hauser has marshaled a persuasive defense for the claim that much complex cognitive analysis of the situation must take place prior to making the subtle and fine grained moral discrimination that people actually make. Moreover, since these discriminations are made quickly and people typically cannot give a convincing account of the considerations involved, most of the mental processing involved must be unconscious, much as it is when we make grammatical judgments. We are, however, not persuaded by Hauser's contention that emotions enter the process only *after* moral judgments are made. The argument that Hauser offers for this aspect of his theory is indirect and far from conclusive, and the Wheatley and Haidt experiment provides some impressive evidence that, sometimes at least, emotions come first and moral judgments follow.

#### 4. Conclusion

The use of empirical methods to explore traditional questions in moral theory is still very much in its infancy, and there is a great deal yet to be learned. In this article we have reviewed just a small sampling of work in this area. Our aim was to illustrate how this work can shed important light on areas of inquiry traditionally of interest to philosophers. At the very least, we hope we have said enough to convince you that it would be intellectually irresponsible for philosophers interested in these issues to ignore the burgeoning, and fascinating, empirical literature in moral psychology.

## Works Cited

- Boyd, R. (2002). "Scientific Realism," *The Stanford Encyclopedia of Philosophy (Summer 2002 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2002/entries/scientific-realism/>>.
- Doris, J. & Stich, S. (2005). "As a Matter of Fact: Empirical Perspectives on Ethics," in F. Jackson and M. Smith, eds., *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press.
- Doris, J. & Stich, S. (2006). "Moral Psychology: Empirical Approaches," *The Stanford Encyclopedia of Philosophy (Summer 2006 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2006/entries/moral-psych-emp/>>.
- Gert, B. (2005). "The Definition of Morality," *The Stanford Encyclopedia of Philosophy (Fall 2005 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2005/entries/morality-definition/>>.
- 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*, Vol. 293, Sept. 14, 2001, 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., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613-628.
- Hauser, M. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: HarperCollins.
- Kelly, D. & Stich, S. (forthcoming). "Two Theories About the Cognitive Architecture Underlying Morality," in P. Carruthers, S. Laurence & S. Stich, eds., *Innateness and the Structure of the Mind*, Vol. III, *Foundations and the Future* (New York: Oxford University Press).
- Kelly, D., Stich, S. Haley, K., Eng, S. Fessler, D. (2007). "Harm, Affect and the Moral / Conventional Distinction," *Mind and Language*, April 2007.
- MacIntyre, A. (1957). "What Morality Is Not," *Philosophy*, 32, 325-335.
- Mikhail, J. (2000). *Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'A theory of justice.'* Unpublished Ph.D. thesis, Cornell University, Ithica, NY.
- Mikhail, J, 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, CA.

Nichols, S. (2002). Norms with feeling: Toward a psychological account of moral judgment. *Cognition*, 84:223-236.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Taylor, P. (1978). "On Taking the Moral Point of View," *Midwest Studies in Philosophy*, v. 3: *Studies in Ethical Theory*, Pp. 35-61.

Turiel, E. (1979). Distinct conceptual and developmental domains: Social convention and morality. In *Nebraska Symposium on Motivation, 1977: Social Cognitive Development*, 25:77-116, eds. Howe, H. & Keasey, C. Lincoln: University of Nebraska Press, 1979.

Turiel, E. (1983). *The Development of Social Knowledge*. Cambridge: Cambridge University Press.

Urmson, J. (1968). *The Emotive Theory of Ethics*, London: Hutchinson.

Unger, P. (1996). *Living High and Letting Die*. New York: Oxford University Press.

Wallace, G. & Walker, A., eds. (1970). *The Definition of Morality*, London: Methuen & Co.

Wheatley, T., & Haidt, J. (2005). "Hypnotically induced disgust makes moral judgments more severe." *Psychological Science*, 16, 780-784.

### Suggested Readings

Doris, J. & Stich, S. (2005). "As a Matter of Fact: Empirical Perspectives on Ethics," in F. Jackson and M. Smith, eds., *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press.

Haidt, J. (2001). "The emotional dog and its 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.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Sripada, C. & Stich, S. (2007). "A Framework for the Psychology of Norms," in P. Carruthers, S. Laurence & S. Stich, eds., *The Innate Mind: Culture and Cognition*. Oxford: Oxford University Press.