Making Psychology Normatively Significant

Regina A. Rini

Abstract The debate between proponents and opponents of a role for empirical psychology in ethical theory seems to be deadlocked. This paper aims to clarify the terms of that debate, and to defend a principled middle position. I argue against extreme views, which see empirical psychology either as irrelevant to, or as wholly displacing, reflective moral inquiry. Instead, I argue that moral theorists of all stripes are committed to a certain conception of moral thought—as aimed at abstracting away from individual inclinations and toward interpersonal norms—and that this conception tells against both extremes. Since we cannot always know introspectively whether our particular moral judgments achieve this interpersonal standard, we must seek the sort of self-knowledge offered by empirical psychology. Yet reflective assessment of this new information remains a matter of substantive normative theorizing, rather than an immediate consequence of empirical findings themselves.

Keywords Cognitive science of ethics · Moral methodology · Moral psychology · Normative abstraction · Intuitions

A pattern has emerged in discussions about the relationship between moral philosophy and moral psychology. Someone points to an interesting new empirical finding, showing that moral judgments are conditioned by some peculiar psychological factor, and uses this finding to indict or support a normative theory. Then someone else points out a flaw in the logic of this argument: the

R. A. Rini (✉)
University of Oxford, Suite 8, Littlegate House, St. Ebbe's Street, Oxford OX1 1PT, UK
e-mail: regina.rini@philosophy.ox.ac.uk

empirical maneuvers have been insufficiently attentive to important normative concepts.[1] Next: repeat with a new interesting empirical finding. This paper is an attempt to make sense of this pattern, to understand the merit there is in each side of an increasingly deadlocked exchange. More ambitiously, I will try to provide a conceptual framework for understanding the relevance of empirical psychology to a priori moral philosophy.

My investigation proceeds in three parts. Parts 1 and 2 combine to form an argument establishing that it is an error for moral philosophers to insist upon the irrelevance of psychological findings to their project. In Part 1 I identify a particular methodological commitment, which I call *normative abstraction*, that I take to be common to moral inquiry. In Part 2 I argue that this commitment can only be seriously upheld with attention to the sorts of self-knowledge provided to us by empirical psychology. However, in Part 3 I attempt to draw a limit around the impact of the preceding argument; I claim that the sorts of trouble raised by psychology can still only be settled by distinctively philosophical reflection. I am not here arguing for any particular side in these disputes—but I do hope to provide some clarity.

1 Moral Thought as Aimed at Normative Abstraction

In *The Last Word*, Thomas Nagel writes, "someone who abandons or qualifies his basic methods of moral reasoning on historical or anthropological grounds alone is nearly as irrational as someone who abandons a mathematical belief on other than mathematical grounds. … [M]oral considerations occupy a position in the system of human thought that makes it illegitimate to subordinate them completely to anything else. (Nagel 1997, 105)" Nagel is arguing for a view that seems difficult to deny: no matter what facts we might discover about the causal antecedents of our moral judgments, only moral reasoning itself can ultimately evaluate them. When I make a moral judgment, it appears to me that I have a specifically *moral* reason to reach some conclusion, the sort of reason responding to a question like, "What should one do in these circumstances?" No claim about the causal antecedents of my judgment seems responsive to this question, and so it is hard to understand how such causal information could justify changing my answer.

In this section I will begin laying out an argument against that very plausible view. There are actually a number of plausible views in the area, so it will help to keep a particular one in mind. I will call this the *Autonomy Thesis*. The term comes from Nagel (1978) and Fried (1978) who argue that ethical theory is *autonomous* of other forms of intellectual inquiry in that its fundamental deliberations do not

---

[1] Examples of the first sort, some of which are discussed in more detail below, include Baron (1995), Horowitz (1998), Sunstein (2005), and Greene (2008). Examples of the second sort include Kamm (1998), van Roojen (1999), Cullity (2006), Berker (2009), and Kahane and Shackel (2010). For good overviews of these debates, see Appiah (2008), Levy (2009), and the essays in Sinnott-Armstrong (2008a).

depend on input from other subjects, especially empirical ones.[2] The fundamental idea is this: empirical disciplines aim at description, explanation, causal theory. Moral philosophy aims at prescription, justification, normative theory. These are different businesses; mixing them is a conceptual error, resulting in confusion or worse. As a conceptual matter, no information about how people *actually* do engage in normative thought can make a substantial difference to how they *should* do so. Therefore, says the Autonomy Thesis, moral philosophy can and should simply ignore such empirical discoveries.[3]

I will build to my argument against the Autonomy Thesis over this section and the next. I need to begin by making some remarks about the nature of prescription. Prescription can, of course, be as simple as *ordering* that something be done, but of course philosophers are interested in when and how prescriptive claims are *justified*. I will therefore this discussion on moral justification, trying to draw out one of its central features. I will then claim, in the next section, that reflecting on this feature exposes a problem for the Autonomy Thesis. This first part will be somewhat imprecise and hand-waving; I will have to ask the reader's patience, in hopes that the point will be clearer once the complete argument is on the table.

Consider a simple example of demand for moral justification. Suppose that I happen to value watching television coverage of the Westminster Kennel Club annual dog show, so much so that I am strongly motivated to do this rather than a range of other possible activities. Suppose, in particular, that in succeeding years I choose to watch the live broadcast rather than attend, respectively, my friend's monthly piano recital, my former student's thesis defense, and finally my grandmother's funeral. If my friend, my student, and my relatives later come to me and demand that I justify my absence at these important events, what can I say in response?

Clearly it is not enough to say "I wanted to watch the dog show". And that response is inadequate not just because my interlocutors will *reject* it. In fact, it will strike them as a non sequitur. Everyone already knows that I wanted to watch the dog show, and they did not take this mere fact as justificatory; they took themselves to be asking a further question. To count as offering a justification in response, I must do more than report my own personal preferences.[4] I must attempt to recast

---

[2] This view is most clearly present in Immanuel Kant, who describes an empirical approach to fundamental moral principles as a "base way of thinking", "disadvantageous to the purity of morals themselves", and "a bastard patched together from limbs of quite diverse ancestry". (Kant 1785/2002, 43–44). Variations on this theme also turn up in Held (1996), Dworkin (1996), and Cohen (2003).

[3] An important qualification: of course there are *some* ways in which empirical information matters to moral theory, even following the Autonomy Thesis. Specifically, empirical information can be crucial in *applying* a moral theory to actual decisions, especially if the moral theory treats individual interests and preferences as input to a decision procedure. For instance, a utilitarian normative theory advising us to maximize individual happiness cannot be implemented without empirical information about what actually makes people happy. Set that sort of application-relevance aside for the moment; what is at issue here is the philosophical relevance of empirical discoveries regarding *how we engage in moral thinking*, and especially the origins of our moral intuitions.

[4] To be clear: there certainly are circumstances in which my mere personal preference *would* be enough to justify my choice. If it is a holiday and nothing else of importance is going on, then of course the mere fact that I want to watch the dog show will count as justifying doing so. However, as the example in the text is meant to demonstrate, mere personal preference falls far short of justification when I am being

this preference in terms that they will recognize as aimed at demonstrating that my preference to watch the dog show meets some interpersonal standard for prioritizing values. I might, for instance, point out that my fondness for the dog show is no passing fancy—I have structured my life around my dedication to dog shows and their role in the culture. Watching the dog show, I say, plays a central role in my conception of a rich, fulfilling life: the same sort of role, perhaps, occupied in *your* life by playing the piano, writing your thesis, or honoring your deceased relatives.

To count as engaged in justification, I need not succeed in *convincing* anyone that I have met this interpersonal standard; my audience might very well decide that the claims I have put forward are not comparable to their own values in the way I have claimed. 'Justification', as I am using it here, pertains to a particular sort of communicative process; it is not a success-word describing certain outcomes of that process.[5] For me to be engaged in justification, my interlocutors must at least be able to recognize that I have *attempted* to respond to their demand for a justification, in a way that I would not had I merely baldly stated my preference for the dog show.

One mark of a moral justification, then, is that it is formulated to appeal to some common standard, shared between the individual providing the justification and an appropriate community (or perhaps a hypothetical community of idealized agents, or diachronic time-slices of the same individual).[6]

But now we can see some complexities around the practice of moral justification. Suppose that I am trying to offer a moral justification. I am only one individual, and I have my own particular set of interests, preferences, and ways of regarding the social world. How do I *know* which of these are entirely particular to me, and which might be relevant to a common standard? Consider the example I introduced a moment ago. There I tried to justify prioritizing the dog show because of the crucial role it plays in my overall conception of a fulfilling life, akin to other people's central projects. In framing the situation this way, I acknowledge that, of course, not everyone holds dog shows in such high esteem as I do. I certainly do not expect everyone to agree with or even understand my preference itself. What I am hoping instead is that others will recognize this preference as being *of the type* of personal preference that one is justified in placing above being present at others' important moments. My search for justification necessarily involves an attempt to get outside my own preferences, to make my choice join with standards which might be endorsed from a perspective other than my own.

Footnote 4 continued

asked why I have failed to fulfil an apparent obligation. The plausibility of a given consideration in an attempted justification will depend on the background assumptions of those demanding justification. Thanks to an anonymous reviewer for *The Journal of Ethics* pressing this point.

[5] One might use 'justification' in a different way, where it *would* be a success-word. One might think that only *successful* efforts to make my preferences acceptable to others (or to some more objective standard) count as 'justifications'. In that case, the non-success-word version of 'justification' used in my text should be read as 'attempted justification'. Nothing really turns on how we elect to use this word, but it is helpful to make note of the fact that my non-success-word use here may differ from the success-word use in other discussions, and hopefully thereby avoid confusion. Thanks to an anonymous reviewer for *The Journal of Ethics* suggesting this clarification.

[6] Prominent views in this vicinity can be found in Smith (1994), Korsgaard (1996), Scanlon (1999), and Darwall (2006).

I have drawn out this feature of moral *justification*, and now I want to suggest that it is distinctive of moral thinking in general. Some particular set of circumstances—some preference, some particular evaluative response to a situation, some particular ordering of values—must be analyzed for features which will bear extraction and generalization, such that various individuals with deeply different viewpoints may attempt a common application of standards. The shift to moral thought, then, is (in large part) a move to a perspective from which actions and preferences are brought under *abstracted* consideration. To engage in characteristically moral judgment is in part to acknowledge that the most particular aspects of one's present attitudes toward the social world must be temporarily suspended. So, writes Hume in his *Enquiry Concerning the Principles of Morals*:

> When a man denominates another his ENEMY, his RIVAL, his ANTAGO-NIST, his ADVERSARY, he is understood to speak the language of self-love, and to express sentiments, peculiar to himself, and arising from his particular circumstances and situation. But when he bestows on any man the epithets of VICIOUS or ODIOUS or DEPRAVED, he then speaks another language, and expresses sentiments, in which he expects all his audience are to concur with him. He must here, therefore, depart from his private and particular situation, and must choose a point of view, common to him with others; he must move some universal principle of the human frame, and touch a string to which all mankind have an accord and symphony. (Hume 1777/1912, section 9, part 1)

Moral thinking, I claim, involves a process that I will call *normative abstraction*. The central idea can be identified in western moral philosophy throughout history and across normative traditions. Immanuel Kant writes that we can only uncover practical laws if we "abstract from the personal differences between rational beings, as likewise from every content of their private ends…" (Kant 1785/2002, 51). Henry Sidgwick writes that "even when a moral judgment relates primarily to some particular action we commonly regard it as applicable to any other action belonging to a certain definable class: so that the moral truth apprehended is implicitly conceived to be intrinsically universal, though particular in our first apprehension of it."[7] John Rawls' (1971) original position is explicitly a device aimed at evaluation of norms from outside a determinate personal perspective. R. M. Hare (1963) claims that the semantics of moral terms entail an implicit commitment to *universalization* in all genuine moral assertions. This is a laundry list of context-less extractions, of course, and these philosophers clearly intended somewhat different things from one another, but all seem to be in this vicinity. Deep in the dominant strand of ethical theory sits a commitment to moral thought as a particular sort of normative abstraction.[8]

---

[7] Sidgwick (1874/1962, 34). He later refers back to this discussion and clarifies: "If therefore I judge any action to be right for myself, I implicitly judge it to be right for any other person whose nature and circumstances do not differ from my own in some important respects." (209) Some process of abstraction seems necessary to determine which "certain definable class" any particular action belongs to, or in what "important respects" two particular actions may or may not be said to differ.

[8] There are, of course, important exceptions. Some people, particularly followers of Aristotelian virtue theory, are likely to reject an account of ethics divorced from particular circumstances and socially embedded contexts. Existentialist philosophers have been especially concerned to emphasize the

Normative abstraction involves determining which of my values, preferences, and the like are merely *idiosyncratic*: those which are particular to me and cannot, on the surface, appeal to any sort of interpersonal standard. I then aim to *abstract away* from these idiosyncrasies, toward a generalized formulation.

Importantly, the claim here is *not* simply that moral thinking incorporates *impartiality*, in the sense of treating everyone's interests equally. Consider Scheffler's (1994) agent-centered prerogatives, which allow agents to weigh their own interests more greatly than the interests of others. Agent-centered prerogatives clearly do not require impartiality, but they *do* require normative abstraction. For, on Scheffler's account, it is permissible for *each* agent to so weigh her or his own prerogatives. And *that* normative claim is itself an abstracted claim; it applies to all agents, and its validity is not supposed to depend in any way on the personal idiosyncrasies of one who considers it. An agent is justified in preferentially weighing her own interests only to the extent that she recognizes that others may permissibly do the same. We might say that normative abstraction is a *formal* criterion on moral judgment, while impartiality is a substantive moral commitment; the two may come apart.

I think that what I have just described is a good-enough characterization of a central aspect of moral thinking; moral thinking is in part a matter of abstracting away from the individual thinker's present personal perspective. Notice, however, that the feasibility of moral thinking then depends upon our ability to identify which aspects of our perspectives actually *are* so idiosyncratic as to require removal in abstraction. Further, it depends upon the assumption that the mental operations we take to constitute abstraction actually *do* strip away personal idiosyncrasies—and not covertly bury them under pleasantly impersonal confabulation. In the next section, I will aim to show that this latter assumption is questionable.

2 Nonconscious Idiosyncrasy and the Failure of the Autonomy Thesis

Had you been a research subject at the University of British Columbia recently, you might have experienced the following (see Proulx and Heine 2008). The experimenter seats you at a table and gives you some forms to complete. While she putters around the lab a bit, you write responses to questions about your entertainment preferences. Then the experimenter returns and gives you a new task: you are to read a hypothetical story about a woman arrested on charges of prostitution and imagine being the judge who must set bond for this woman. You write your answer and turn in your forms, but now the experimenter asks some very unexpected questions. Did you notice anything odd about the experiment? Did you notice anything odd about the *experimenter*?

Footnote 8 continued

ineradicable centrality of personal standards in moral determination. And moral particularists (Dancy 2004, Hooker and Little 2000) are obviously unlikely to agree that moral deliberation necessarily involves abstraction, at least not without qualification.

If you are like 90 % of subjects in this experimental condition, you did not notice anything. In particular, you did not notice that the experimenter now asking you questions is not the same experimenter who sat you down at the table a few minutes earlier! When she went to the cabinet and stepped very briefly behind its large opaque door, she switched places with another experimenter—fairly similar looking, and wearing identical clothes, but very much a different human being. But you did not notice.[9]

Or did you? In fact, if you are like most other subjects in this condition, the experimenter switch had *some* impact on you, just not at a conscious level.[10] It turns out that the bond figure you set for the hypothetical prostitution suspect is probably significantly higher (on average, almost 50 % higher) than that chosen by subjects in the control condition, who interacted with the same experimenter throughout. Apparently, having been exposed to this disconcerting person replacement—even if you are not consciously aware that you have witnessed it—inclines you to assign more severe penalties for an unrelated social infraction.[11]

I describe this study to serve as a particularly illustrative example of an increasingly well-documented phenomenon. Our moral judgments are evidently affected by various psychological processes to which we have no introspective access—and which we would likely repudiate were they to become apparent. Consider the situation of a typical subject in this study. He is asked to consider the case of the hypothetical prostitution suspect and then to make an evaluative judgment about her. More particularly, he is asked to i*magine being a judge*—a position marked by the expectation of impersonal deliberation—and to make an evaluative judgment from *within that perspective*. Significantly, then, the subject is

---

[9] You should not feel too badly about not noticing the switch though, since so few people do. Indeed, this is a well-established result in the literature on "change blindness" in perception research (Simons and Levin 1998, Levin et al. 2002).

[10] To avoid any confusion: here my use of the word "nonconscious" (and similar locutions) is intended to deny the presence of what Ned Block (1995, 2007) calls *access* consciousness, the availability of a particular representation to global mental operations, especially verbal report.

[11] One might object to interpreting this study as one bearing on moral judgment. Subjects might have regarded setting a bond figure *not* as a form of moral sanction, but merely as an attempt to deter suspect flight—and the experimenter-swap effect may have unsettled their confidence in predicting the suspect's reliability, rather than affecting moral judgment directly. (Thanks to an anonymous reviewer for *The Journal of Ethics* this objection.) I think there is some merit in this point, but it is worth noting possible replies. First, unlike in other legal regimes, the criminal law of Canada (where this study took place) explicitly *does* require bond-setting judges to consider factors other than flight risk, such as whether granting bond would "maintain confidence in the administration of justice" in a manner fitting "the gravity of the offense". (*Criminal Code of Canada*, section 515.10(c). See http://laws-lois.justice.gc.ca/eng/acts/C-46/section-515.html) Second, the authors of the study themselves state that the bond figure is intended as a measure of whether "people are motivated to maintain their cultural worldview and will seek to punish individuals who act in ways that are inconsistent with that worldview. (Proulx and Heine 2008, 1296)" They claim that the measure has been utilized for this purpose in a series of related empirical studies. In their interpretation, the findings of this study are support for the psychological concept of "compensatory affirmation", whereby the subject responds to an irresolvable disruption in one psychological schema (person consistency) by temporarily heightening expression of some unrelated but manageable schema (enforcement of social norms). So, while the objection does have some merit, addressing it fully would require working against a significant body of legal and psychological literature. In any case, I focus on this study as an illustration; the general point about empirically-demonstrated nonconscious factors in moral judgment can be made with other studies (mentioned below) if one does not trust this one.

not being asked simply to register a personal opinion, but instead to (imaginatively) render a verdict on behalf of the entire community.

You already know the rest of the story. Despite instructions to deliberate from this detached perspective, subjects' responses are affected by highly personal aspects of their immediate situation; subjects in the switched-experimenter condition are willing to endorse applying a far larger social sanction to the hypothetical prostitution suspect than are control subjects. Of course, this does not mean that they deliberately failed to follow instructions. So far as they know, they *did* successfully adopt a detached perspective. The switched-experimenter manipulation made its contribution at some nonconscious, subpersonal level, as demonstrated by these subjects' inability to consciously detect it even when prompted.

As I have suggested, these results are by no means unique (apart from the instruction to imagine being a judge, a feature with particular salience to this discussion). In a range of contexts and designs, experimenters have shown that subjects' moral evaluation can be covertly manipulated with remarkable ease. People deliver harsher moral verdicts while sitting at a filthy desk rather than a clean one (Schnall et al. 2008). Their judgments vary after watching a funny movie (Valdesolo and DeSteno 2006), or consuming a sugary beverage (Eskine et al. 2011). What is common to each is the empirically (but not introspectively) detectable influence of some contextual factor on subjects' moral verdicts and behavior, where on reflective consideration we cannot endorse the relevant contextual factor as a basis for moral differentiation. In other words, these experiments show that our moral decisions are implicitly shaped by considerations which we ourselves do not regard as morally relevant.

These studies suggest that our spontaneous moral judgments—call them moral *intuitions*, if you like—are the output of some set of subpersonal, nonconscious psychological processes. A very influential view in contemporary psychology suggests that our minds contain 'dual processes': a rapid, nonconscious, heuristic-based system for making quick and good-enough judgments, and a slow, conscious reasoning system that trades efficiency for greater carefulness. On this psychological model, efficient nonconscious processes are usually understood to be *cognitively impenetrable* to conscious reasoning: we cannot consciously inspect the manner in which our nonconsciously-generated judgments arise.[12] Hence the gap between factors apparently influencing our nonconscious thinking and what we can consciously endorse upon reflection.[13]

---

[12] Dual-process models are described at length in Stanovich and West (2000) and Kahneman (2002). The idea of cognitively impenetrable processes owes much to Fodor (1983).

[13] The gap itself is a problem if we accept what Christine Korsgaard calls the *transparency condition* for moral theory: "A normative moral theory must be one that allows us to act in the full light of knowledge of what morality is and why we are susceptible to its influences, and at the same time to believe that our actions are justified and make sense." (Korsgaard 1996, 17) Given the transparency condition, the possibility of nonconscious influence upon moral thinking provides a direct argument against the Autonomy Thesis; transparency requires that we understand, through empirical means if necessary, why we value what we value.

More to the point, the studies described above suggest that our moral thinking sometimes fails to meet the demands of normative abstraction. Our moral judgments are apparently sensitive to *idiosyncratic* factors, which cannot plausibly appear as the basis of an interpersonal normative standard. The fact that I happened to be drinking a sugary beverage or in the presence of a swapped experimenter could not possibly appear as justificatory to other thinkers (or to myself, upon reflection).

Of course, merely observing that we contain idiosyncrasies affecting our moral thought is not itself a threat. If we could monitor our nonconscious systems and somehow identify and abstract *away from* their idiosyncrasies, then we would accomplish precisely what is required in adopting a moral stance. The trouble, of course, is that idiosyncrasies buried in our subpersonal, nonconscious processes are not available for identification and abstraction, because those processes are impenetrable to conscious introspection. I cannot figure out how my subpersonal, nonconscious processes work simply by introspecting upon my thought in the relevant domain. The subjects in the studies above were not aware of the factors influencing their moral judgments. So we are not in a position to introspectively isolate and abstract away from these factors.

Worse yet, even when we *think* that we have achieved normative abstraction, we may only *erroneously* conclude that we have succeeded. Consider one extremely suggestive study, in which the psychologists Thalia Wheatley and Jonathan Haidt used hypnosis to *create* in their subjects an introspectively undetected evaluative reaction.[14] Subjects were given a posthypnotic suggestion to feel "a brief pang of disgust… a sickening feeling in the stomach" whenever they heard either the word "take" or the word "often" (Wheatley and Haidt 2005). After hypnosis (and a check to eliminate those who were aware of the hypnotic suggestion), subjects were asked to rate the moral wrongness of various verbally described actions, in each of which the word "take" or "often" appeared unobtrusively. (For instance, a Congressman is described as willing "to take bribes" or as "often bribed".) Compared to subjects who had not received the relevant posthypnotic suggestion, the hypnotized subjects gave significantly higher wrongness ratings.

Most important for present purposes is the behavior of subjects asked to verbally account for their evaluations. In one particularly instructive instance, subjects given the posthypnotic suggestion displayed a strange willingness to attribute some degree of moral wrongness to the seemingly innocuous, even laudable, behavior of a student council president who "[tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion." While many subjects were unable to articulate any justification for their responses, others gestured at rather implausible claims: e.g. the student council president was "up to something" or "a popularity seeking snob".

I interpret this as follows: asked to give an interpersonal justification of their negative evaluation, subjects reasonably sought out something responsive to the demands of abstraction. The student council president's being biased or unfair, especially as he holds a position responsible for neutral mediation, would surely be something condemnable. Yet clearly these subjects are making a mistake. The

---

[14] This study is also discussed, at similar length and with similar aim, in Appiah (2008, 86–87).

student in the story *is not* "up to something", and there seem no grounds to accuse him of popularity-seeking snobbery. Subjects who had not received the posthypnotic suggestion did not make claims of this sort.

This instance, I suggest, like the experimenter-swapping study, allows us to closely see the danger of failed abstraction. These subjects aimed at abstracting away from their personal reaction (the "pang of disgust" induced by hypnosis) by supposing that this reaction *must* be explained by *some* interpersonal standard, and grasped at the most salient one available. Yet they simply failed to abstract from their subjective reaction—instead they unwittingly buried it in seemingly interpersonal terms. We can see that this happened, because we know the full story behind their reaction. But *they* cannot know this, at least not introspectively. And once they do learn this—when they are released from the posthypnotic suggestion and debriefed—they acquire good reason to suspect the justificatory status of their earlier ruminations.[15]

While we cannot hope to introspectively detect the idiosyncrasies infecting our moral thought, the techniques of experimental psychology, neuroscience, biology, and sociology (and others) provide indirect ways of discovering them. We can only successfully abstract away from our individual situations if we first identify just what those situations are (including our subpersonal, nonconscious situations), and *this* requires the aid of empirical disciplines.[16]

Here, then, is the full argument against the Autonomy Thesis. Abstraction away from personal idiosyncrasies is an essential trait of genuine moral thought. But our attempts to engage in such abstracted thought are vulnerable to idiosyncratic aspects of our subpersonal, nonconscious processes, and these idiosyncrasies are only identifiable with empirical aid. Hence the Autonomy Thesis—holding that moral philosophy may be safely conducted without attention to the psychology underlying our moral deliberations—cannot be correct.

3 The Role of Descriptive Psychology in Moral Theorizing

I have stated quite a bit to establish that empirical psychology must be consulted to achieve one of the central aims of moral theorizing. But I have not yet stated anything about how such consultation is to proceed, or what its limits might be. In this concluding section I turn to those questions, abstractly first, then through examples.

I have argued that empirical investigation allows us to identify psychological factors that influence our moral judgments, yet which we do not reflectively regard as surviving normative abstraction. It is important to note that there are two distinct

---

[15] Importantly, this does not show what Wheatley and Haidt suggest in their discussion—that *all* forms of moral judgment are post hoc or biased. What it does show, however, is that we are very bad at *detecting* when we are biased or engaging in mere post hoc rationalization.

[16] Sinnott-Armstrong (2008b, 75) reaches a very similar conclusion: "Moral intuitionists cannot simply dismiss empirical psychology as irrelevant to their enterprise. They need to find out whether the empirical presuppositions of their normative views are accurate." His argument differs from mine in that it focuses on general epistemic standards of reliability, while mine is grounded in a concern (normative abstraction) particular to the moral domain.

components to this task, and that the role of empirical psychology comes in only one of them. That is, we must (a) discover the psychological factors influencing our moral judgments, and (b) decide among these which are mere idiosyncrasies and which can withstand the move to normative abstraction. Empirical psychology is essential to the first of these, but its relevance to the second is extremely limited.

For example, consider our tendency to value the well-being of family members over that of other people. It does not, of course, take much empirical investigation to uncover this tendency. But suppose that we discover its effects are more pervasive and substantial than we might have thought. Suppose systematic empirical testing reveals that most people are willing to *severely discount* the well-being of other people, relative to that of their own family members. Perhaps we will find that many people would choose to save the life of a single cousin rather than 10,000 unrelated individuals. We might even discover precise algorithmic laws linking genetic similarity and a subject's attitude toward particular moral patients.[17]

Such a precise characterization of the factors influencing this set of moral judgments could only come about through empirical means. But once we have this information, our decision regarding what do about it plainly is not a matter of empirical psychology. We must ask ourselves: does this sort of family-privileging attitude withstand normative abstraction? That is, can we understand our overwhelming preference for preserving kin as something that will appear interpersonally justificatory? Or does it appear now to express an idiosyncratic preference that will need to abstract away from? Answering these questions requires thinking carefully about the nature of interpersonal justification, about what sorts of things are truly normative. This, of course, is precisely a question of moral philosophy, and not very different from how such questions have always been understood by moral philosophers. There has always been a question of weighing familial bonds and obligations to others, at least as far back as Socrates' querying of Euthyphro. Empirical psychology can allow us to explore the mechanisms of this contrast, but it alone is never in a position to settle the questions that arise.

An inquiry of exactly this sort plays out in Peter Singer's book *The Expanding Circle*. There, Singer argues that our preference for providing benefits to our own children—rather than saving the lives of people starving far away—can be best explained by kin selection theory in evolutionary biology. At first, Singer appears to be tempted to claim that kin preferences of this sort are not deserving of moral respect, drawing an analogy to entrenched racial bias (Singer 1981, 33). However, he concludes that the biological drive to favor one's family can be justified on familiarly consequentialist grounds: it is most efficient for the sick and needy to be cared for swiftly by those with strong desires to do so, rather than falling on the disorganized mercy of society at large (36). As this example shows, even someone as ready to debunk commonly-held views as is Singer can recognize a psychological basis for our moral judgments, yet still proceed to evaluate those judgments on

---

[17] There is an enormous literature on the mathematics of kin-selection and its expression in animal and human behaviour, starting from Hamilton (1964). The details of any particular proposal don't matter to the present point about methodology, although they might in coming to a substantive conclusion on our obligations to kin.

clearly philosophical grounds.[18] Psychology is an aid to reflective self-understanding, but it does not decide these matters.

Sometimes, though, we seem to lose track of the essential role of moral theory in analyzing psychological findings.[19] I suggest that this tendency itself can be given a psychological explanation! Our move toward normative abstraction in evaluating psychological research is so swift, so *automatic*, they we do not even notice that we are doing it. We learn of some new experiment in which subjects have been induced to make silly judgments by some subtle environmental variable (dirty desks, transmogrifying scientists, etc.) and *immediately* we are amused by the silliness of their response. We do not need to think very hard about it. We are not usually tempted to think that the experimental manipulation has caused subjects to deliver the *correct* moral verdict.[20] We are confident instead that they have *not* done so. But this immediate judgment of ours does depend on our possession and use of a background normative theory, or at least an implicit sense of what sorts of things are grounds for interpersonal justification. And our use of that theory is so quick, and so effortless, that it is easy to simply forget that our background normative views had any role to play at all. We notice the surprising finding and the clever psychological technique, and we are startled by this unexpected discovery about our own minds. But we do not really notice the evaluative contribution of our own normative judgment—in a sense, we *look right through it*. It is, then, easy for it to *seem* as if our discovery of this new challenge to moral judgment arises directly and solely from psychology, as if the problem could be read right off psychological data. But it cannot.[21]

Because we are prone to overlooking the role of our own normative judgment in our reaction to psychological findings, it is also easy to fail to appreciate the need to embed this reaction in a fleshed out moral theory. This is an important point,

---

[18] My discussion of Singer here is intended only as an *example* of how one's reaction to psychological explanation of moral intuition might change through prolonged moral reflection. I take no position on the details of Singer's reasoning. In particular, my conception of normative abstraction does not necessarily require abstracting to quite the degree that Singer does. (He regularly suggests that moral justification can ultimately be secured only from Sidgwick's "point of view of the universe"). There are grounds for deep philosophical disagreement about the *degree* of abstraction required for moral justification; I do not aim to settle such issues in this paper. (Thanks to an anonymous reviewer for *The Journal of Ethics* pressing me to clarify this point).

[19] This tendency seems to have particular traction in popular accounts of the relation between psychology and moral philosophy; see David Brooks, "The End of Philosophy" (*The New York Times*, April 6, 2009) and *The Economist,* "Moral thinking: biology invades a field philosophers thought was safely theirs" (February 21, 2008).

[20] The exceptions to this rule—when a manipulation *does* seem to generate the correct response—confirm the underlying point. For instance, Caruso and Gino (2011) claim to show that subjects behave more ethically after deliberating with their eyes closed. Their closed-eyed subjects were more generous and indicated less willingness to engage in dishonest behavior. Of course, characterizing these subject behaviors as "ethical" must depend on a background moral theory about generosity and honesty!

[21] F. Nietzsche appreciated early the need to distinguish our reaction at learning a causal story about moral judgments from the story itself, and the need to take this reaction up in an appropriately reflective spirit: "The inquiry into the *origin of our evaluations* and tables of the good is in absolutely no way identical with a critique of them, as is so often believed: even though the insight into some *pudendo origo* certainly brings with it *a feeling* of a dimunition in value of the thing that originated thus and prepares the way to a critical mood and attitude toward it."(Nietzsche 1901/1967, section 254).

because once we do begin to explicitly consider the role of moral theory in generating our reaction to new psychological findings, we may come to re-evaluate the reaction itself. That is, if we attend to the (implicit) employment of moral theory in our response to psychological findings, and try to work through how to make this role explicit, we might find that the psychological findings are not quite as worrisome as they initially seemed. While it may at first seem that some nonconscious psychological process is idiosyncratic, careful thought might show it to actually be able to support normative abstraction. This, I suggest, describes Singer's ruminations on our biologically-grounded preferences for kin.

I think that a failure to make explicit the contribution of moral theory has hampered many discussions of moral psychology in extant literature. I will conclude by briefly discussing two representative examples, showing how they tempt us to make this mistake.

Consider the purported role of cognitive heuristics and biases in moral judgment. Tamara Horowitz (1998) argues that our confidence in a familiar principle of normative ethics, the Doctrine of Doing and Allowing (see, e.g. Quinn 1989), can be best explained as the effect of a set of psychological tendencies known as framing and base rate neglect (Kahneman and Tversky 1979). Very roughly, Horowitz's claim is that our differential attitudes towards doing harm to a person and allowing harm to befall a person are simply a particular instance of our general tendency to prefer safety when a choice is framed as a gain and to prefer risk when a choice is framed as a loss—even when the choices are equivalent in all other respects. It is not entirely clear what normative lessons Horowitz intends to draw from this claim, but she does state that intuitions supporting the traditional Doctrine of Doing and Allowing are here shown to be responsive to a "morally irrelevant" factor, and therefore "are not moral intuitions at all." (381) Presumably this must cast doubt on the Doctrine.

Horowitz writes as if this conclusion simply falls out of the psychology, but clearly there is a suppressed normative premise here. Her grounds for declaring the moral irrelevance of these psychological factors cannot come from anywhere but some conception of what factors *would* be morally relevant—which is to say, some moral theory. The appropriate response, then, is not to move immediately toward the conclusion that psychological findings have undermined the Doctrine of Doing and Allowing (as, e.g. Sunstein (2005) does) but instead to take a step back and ask, from the perspective of normative theory, whether our commitment to the Doing and Allowing distinction can survive this new information: Is this psychological characterization of our judgments somehow still compatible with how we conceive of the role of these judgments in motivating us to respond to certain sorts of actions (or omissions)? Indeed, Kamm (1998) attempts precisely this sort of inquiry, and concludes that—in some ways, at least—it is.[22]

For another example, consider some extremely influential research by philosopher-neuroscientist Joshua D. Greene (Greene et al. 2001, 2004; Greene 2008), and certain subsequent philosophical appropriations thereof. This research

---

[22] For more on normative conclusions drawn from the biases and heuristics literature, see Kahneman (1994), van Roojen (1999), Sinnott-Armstrong (2008b), and Gigerenzer (2008).

purports to show that when subjects engage in characteristically deontic judgments—such as when subjects indicate rejection of intentionally killing one innocent to save five—their brains exhibit enhanced activity in regions correlated with emotional excitement. Further, according to Greene, when subjects engage in characteristically consequentialist judgments—as when they approve of killing one to save five—their brains exhibit enhanced activity in regions correlated with cognitive conflict and control, which Greene describes as "rational" areas of the brain.[23]

Greene argues that these data reveal that deontic judgments originate in normatively suspect psychological processes. In contrast, "The only way to reach a distinctively consequentialist judgment… is to actually go through the consequentialist, cost-benefit reasoning using one's 'cognitive' faculties, the ones based in the dorsolateral prefrontal cortex." (Greene 2008, 65) In other words, we may trust consequentialism because it arises from rational brain areas, while deontology does not. Peter Singer makes a similar point from a Darwinian angle: the consequentialist verdict that we may sacrifice one innocent to save five "is different… It does not seem to be one that is the outcome of our evolutionary past." (Singer 2005, 350) Again, we should disregard moral evaluations stemming from our species' peculiar developmental history, in favor of those that do not. According to Greene and Singer, such contingent evolutionary happenstance is just what we aim to abstract away from when we adopt a normative perspective.

As many readers have noted, this argument appears to assume some very strong premises, which are not always explicitly defended. There must be some substantive claim about the normative status of brain areas or evolutionary trajectories; it is not enough to draw a connection between certain normative theories and these brain areas or evolutionary histories and then gesture disapprovingly at the latter. Unfortunately, Greene and Singer provide relatively little guidance on this point: Greene states that deontic judgments uniquely respond to "up-close-and-personal harms", which means they are triggered by a "non-moral feature of our evolutionary history" (70). Why is the "up-close-and-personal" feature of a harm "non-moral"? Greene does not really answer this question.

What seems more likely is that his judgment stems from a prior normative evaluation of deontology itself, rather than any new conclusion arising from the neuroscience. Singer in fact moves fairly quickly to asserting exactly these normative judgments as support for the empirical methodology itself. He suggests that he and Greene are right to describe as 'cognitive' the brain area active when subjects endorse killing one innocent to save five, because "the answer these subjects give is, surely, the rational one. The death of one person is a lesser tragedy than the death of five people." (Singer 2005, 350) But if that is obviously the "rational" response, then what difference does it make what the neuroscience finds? Again, it seems clear that any conclusion against deontology will not follow

---

[23] Some have expressed deep *empirical* methodological reservations about this research, especially in how it experimentally operationalizes the key concepts of *deontology*, *consequentialism*, *emotion*, and *reason*. See Berker (2009), Kamm (2009), and Kahane and Shackel (2010). But I will set aside such complaints for now to concentrate on interpretation of the results.

straight-away from the empirical findings; it requires a far more careful normative apparatus than Greene and Singer have supplied.[24]

Some (e.g. Berker (2009)) have concluded from a similar analysis of this example that empirical findings are "insignificant" to normative theory; the important work is done by normative premises, so empirical findings simply disappear from the argument. I think this is an overreaction. After all, if Greene's empirical claims are right, they *do* constitute a jarring discovery for deontologists. Learning that our deontological judgments selectively arise from attention to "up-close-and-personal" harms *does* affect the way we approach these problems, because it comes as a *surprise*: this is not the feature we thought we were responding to in making these judgments. And our initial reaction to this discovery is indeed to be newly suspicious of the moral judgments in question. But, again, we should attend to this suspicion, and attempt to locate its source in a broader normative theory.

The question can be reframed as one about whether selective response to up-close-and-personal harms can support interpersonal justification. It might, for instance, be that this sort of harm demonstrates a particularly objectionable attitude toward others, one absent from other means of causing harm. When we work through the details, deontologists may come to find their overall assessment restored to equanimity; it may turn out that they can reflectively *endorse* their having minds that make them attend to "up-close-and-personal" harms.[25] I do not aim to settle the matter here, only to point out that our ability to appreciate this new framing of the debate *depends* upon empirical psychology—it just will not be resolved by it.[26]

---

[24] Similar points have been raised against Greene and Singer by others, including Berker (2009, 326), Kamm (2009), and Cullity (2006, 127). Greene himself, at least in less formal contexts, falls easily into employing normative premises for his argument. In a perceptive interview conducted by the philosopher Tamler Sommers and with the neuroscientist Liane Young, Greene is challenged with the empirical fact that consequentialist intuitions are *also* correlated with activity in some "emotional" areas of the brain—just not the same emotional areas as are deontic intuitions. Asked to explain why he favors one set of emotion-correlated intuitions over another, Greene does not offer a theory about the relative superiority of certain emotions or brain-areas. Rather, he appeals to typical utilitarian considerations. [See Sommers (2009, pp. 138–141)]. However, in an unpublished paper, Greene (manuscript) has been developing a bit more guidance about the normative superiority of the cognitive systems driving consequentialist intuitions, though the strength of this argument remains to be seen.

[25] I should say that I have been a bit unfair to Singer, who is clearly aware of the need to provide a normative framework. He writes, "Advances in our understanding of ethics do not themselves directly imply any normative conclusions, but they undermine some conceptions of doing ethics which themselves have normative conclusions. Those conceptions of ethics tend to be too respectful of our intuitions. Our better understanding of ethics gives us grounds for being less respectful of them." (Singer 2005, 349) The trouble is that his assertion that deontology is "too respectful of our intuitions" rests on familiar consequentialist grounds, and nothing in *this* exchange seems to advance that dialectic. In a more recent paper, with Katarzyna de Lazari-Radek, Singer provides more detail about the evolutionary debunking of certain moral judgments, once again on Sidgwickian, utilitarian grounds (de Lazari-Radek and Singer 2012).

[26] Intriguingly, C. L. Stevenson—surely Greene's meta-ethical forebear—seems to have anticipated almost precisely this exchange: "If certain of our attitudes are shown to have the same origin as the taboos of savages, we may become disconcerted at the company we are force to keep. After due consideration, of course, we may decide that our attitudes, however they may have originated, are unlike many taboos in that they will retain a former function, or have since acquired new ones. Hence we may insistently preserve them. But in the midst of such considerations we shall have been led to see our attitudes in a

Indeed, the aim of this paper has not been to *settle* any disputes, but to clarify a point of methodology. I have argued that attempts to resist the intrusion of empirical moral psychology into the practice of moral philosophy are destined to be unsuccessful; we must learn to appreciate and effectively use the self-knowledge delivered by these techniques. But this needs to be understood as an evolution of traditional philosophical methods, not their wholesale displacement by the march of unsentimental, objective science. This is a subtle balance, made all the more difficulty by the dizzying stereoscopy of viewing side-by-side our introspective self-conception and the schematic of our minds drawn by empirical psychology. For moral philosophy to be in any way true to its aims, finding such a balance is inescapable. Our challenge is to look simultaneously inward and outward, toward experimentation and toward reflection, and from these parallel perspectives to fashion an informed study both of how we are *and* of how we ought to be.

References

Appiah, Kwame Anthony. 2008. *Experiments in ethics*. Cambridge: Harvard University Press.
Baron, Jonathan. 1995. A psychological view of moral intuition. *Harvard Review of Philosophy* 5: 36–40.
Berker, Selim. 2009. The normative insigificance of neuroscience. *Philosophy & Public Affairs* 37: 293–329.
Block, Ned. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227–247.
Block, Ned. 2007. Consciousness, accessibility, and the mesh between psychology and neural science. *Behavioral and Brain Sciences* 30: 481–499.
Caruso, Eugene M., and Francesca Gino. 2011. Blind ethics: Closing one's eyes polarizes moral judgments and discourages dishonest behavior. *Cognition* 118: 280–285.
Cohen, G.A. 2003. Facts and principles. *Philosophy & Public Affairs* 31: 211–245.
Cullity, Garrett. 2006. As you were? Moral philosophy and the aetiology of moral experience. *Philosophical Explorations* 9(1): 117–132.
Dancy, Jonathan. 2004. *Ethics without principles*. Oxford: Clarendon Press.
Darwall, Stephen. 2006. *The second person standpoint: Respect, morality, and accountability*. Cambridge, MA: Harvard University Press.
de Lazari-Radek, Katarzyna, and Peter Singer. 2012. The objectivity of ethics and the unity of practical reason. *Ethics* 123(1): 9–31.
Dworkin, Ronald. 1996. Objectivity and truth: You'd better believe it. *Philosophy & Public Affairs* 25(2): 87–139.
Eskine, Kendall J., Natalie A. Kacinik, and Jesse J. Prinz. 2011. A bad taste in the mouth: gustatory disgust influences moral judgment. *Psychological Science* 22: 295–299.

Footnote 26 continued

natural setting, and shall be more likely to change them with changing conditions. Hence anyone who wants to change a man's attitudes can prepare the way by a genetic study." (Stevenson 1944, 123–124).

Fried, Charles. 1978. Biology and ethics: Normative implications. In *Morality as a biological phenomenon*, ed. G. Stent, 187–197. Berkeley: University of California Press.

Fodor, Jerry. 1983. *The modularity of mind*. Cambridge: MIT Press.

Gigerenzer, Gerd. 2008. Moral intuition = fast and frugal heuristics? In Sinnott-Armstrong (2008a), 1–26.

Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.

Greene, J.D., L.E. Nystrom, A.D. Engell, J.M. Darley, and J.D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.

Greene, Joshua D. 2008. The secret Joke of Kant's Soul. In *Moral psychology, Vol. 3: The neuroscience of morality: Emotion, disease, and development*, ed. W. Sinnott-Armstrong, 35–80. Cambridge: MIT Press.

Greene, Joshua D. (manuscript). Beyond point-and-shoot morality: Why Cognitive Neuro (Science) Matters for Ethics.

Hamilton, W.D. 1964. The genetical evolution of social behavior. *Journal of Theoretical Biology* 7(1): 1–16.

Hare, R.M. 1963. *Freedom and reason*. Oxford: Oxford University Press.

Held, Virginia. 1996. Whose Agenda? Ethics versus cognitive science. In *Minds and morals: Essays on ethics and cognitive science*, ed. May Friedman, and E. Clark, 69–88. Cambridge, MA: MIT Press.

Hooker, Brad, and Margaret Little. 2000. *Moral particularism*. Oxford: Oxford University Press.

Horowitz, Tamara. 1998. Philosophical intuitions and psychological theory. *Ethics* 108: 367–385.

David, Hume. 1777/1912. *An enquiry concerning the principles of morals*. Available online through Project Gutenberg. http://www.gutenberg.org/files/4320/4320-h/4320-h.htm.

Kahane, Guy, and Nicholas Shackel. 2010. Methodological problems in the neuroscience of moral judgment. *Mind Language* 25(5): 561–582.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2): 263–292.

Kahneman, Daniel. 1994. The cognitive psychology of consequences and moral intuition. The tanner lecture in human values, University of Michigan, Ann Arbor, November 1994. (as cited in Kamm 1998).

Kahneman, Daniel. 2002. Maps of bounded rationality: A perspective on intuitive judgment and choice. In: *Le Prix Nobel*, ed. T. Frangsmyr, 416–499. Nobel Foundation: Stockholm.

Kamm, Frances. 1998. Moral intuitions, cognitive psychology, and the harming-versus-not-aiding distinction. *Ethics* 108(3): 463–488.

Kamm, Frances. 2009. Neuroscience and moral reasoning: A note on recent research. *Philosophy & Public Affairs* 37(4): 330–345.

Kant, Immanuel. 1785/2002. *Groundwork for the Metaphysics of Morals*. Trans. A.W. Wood. New Haven: Yale University Press.

Korsgaard, Christine M. 1996. *The sources of normativity*. Cambridge: Cambridge University Press.

Levin, D.T., et al. 2002. Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology* 93: 289–302.

Levy, Neil. 2009. Empirically informed moral theory: A sketch of the landscape. *Ethical Theory and Moral Practice* 12(1): 3–8.

Nagel, Thomas. 1978. Ethics as an autonomous theoretical subject. In *Morality as a biological phenomenon*, ed. G. Stent, 198–208. Berkeley: University of California Press.

Nagel, Thomas. 1997. *The last word*. Oxford: Oxford University Press.

Nietzsche, Friedrich. 1901/1967. *The will to power*. Trans. W. Kaufmann. New York: Vintage Books.

Proulx, T., and S.J. Heine. 2008. The case of the transmogrifying experimenter: Reaffirmation of moral schema following implicit change detection. *Psychological Science* 19: 1294–1300.

Quinn, Warren. 1989. Actions, intentions, and consequences: The doctrine of doing and allowing. *Philosophical Review* 98(3): 287–312.

Rawls, John. 1971. *A theory of justice*. Cambridge: Harvard University Press.

Scanlon, T.M. 1999. *What we owe to each other*. Cambridge: Harvard University Press.

Scheffler, Samuel. 1994. *The rejection of consequentialism*. Oxford: University Press.

Schnall, S., J. Haidt, G.L. Clore, and A.H. Jordan. 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34: 1096–1109.

Sidgwick, Henry. 1874/1962. *The Methods of Ethics*. Chicago: University of Chicago Press.

Simons, D.J., and D.T. Levin. 1998. Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review* 5: 644–649.

Singer, Peter. 1981. *The expanding circle: Ethics and sociobiology*. Oxford: Oxford University Press.

Singer, Peter. 2005. Ethics and intuitions. *The Journal of Ethics* 9(3/4): 331–352.

Sinnott-Armstrong, Walter. 2008a. *Moral psychology, vol. 2: The cognitive science of morality*. Cambridge, MA: MIT Press.

Sinnott-Armstrong, Walter. 2008b. Framing moral intuitions. In Sinnott-Armstrong (2008a).

Smith, Michael. 1994. *The moral problem*. New York: Wiley-Blackwell.

Sommers, Tamler. 2009. *A very bad wizard: Morality behind the curtain*. New York: McSweeney's.

Stanovich, Keith E., and Richard F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23: 645–726.

Stevenson, Charles L. 1944. *Ethics and language*. New Haven: Yale University Press.

Sunstein, Cass R. 2005. Moral heuristics. *Behavioral and Brain Sciences* 28: 531–573.

van Roojen, Mark. 1999. Reflective moral equilibrium and psychological theory. *Ethics* 109: 846–857.

Valdesolo, Piercarlo, and David DeSteno. 2006. Manipulations of emotional context shape moral judgment. *Psychological Science* 17: 476–477.

Wheatley, Thalia, and Jonathan Haidt. 2005. Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16(10): 780–784.