Can Moral Obligations Be Empirically Discovered?

JESSE J. PRINZ

These are good days for empirically minded philosophy. It is now common to see philosophers cite laboratory studies in support of their theories, and a growing number of philosophers are conducting their own psychological experiments. From a worm's eye perspective, it looks a little bit like there is a methodological revolution taking place. Gone are the days when philosophers had to read psychology journals covered in discrete bindings so as to hide a perverse interest in the latest sexy results. At least that is true in most areas of philosophy. Lately, the methodological revolution has been gaining ground in ethics. Philosophers now routinely site the burgeoning empirical literature on moral judgment. The majority of this work has been descriptive in nature. Psychology and neuroscience are being used to determine what goes on in people's heads when they decide that something is morally right or morally wrong. Traditionally minded ethicists continue to feel insulated from the onslaught of empirical moral psychology. Ethics has long been in the business of telling how we ought to behave, rather than merely describing how we do behave. Such normative claims appear to have a kind of immunity to empirical investigation. Psychology can reveal what we morally detest, but it cannot reveal what we should detest. Or so it would seem. In this chapter, I want to explore this supposition and argue that empirical methods can actually shed light on what we ought to do. Along the way, I will also make some other controversial claims. I will say that morality has an emotional basis, and I will endorse a form of relativism. Of course, such conjectures require much more discussion than I can offer here. My arguments will be preliminary, and I will content myself with having shown how, in principle, one might go about trying to discover obligations

272   *Jesse J. Prinz*

empirically. Along the way, I will indulge in periodic reflections on methodology, which can be accepted quite independently of the contentious philosophical conclusion I draw. If my philosophical conclusions are rejected, I hope they are rejected on empirical grounds.

STEP 1: THE EMOTIONAL BASIS OF MORAL CONCEPTS

Moral judgments, such as the judgment that it is wrong to sleep with your best friend's lover, contain moral concepts, such as the concept WRONG. There is no shortage on controversy about the nature of concepts, but according to the majority view in cognitive science (and increasingly in philosophy) concepts are mental representations: concrete particulars in the head that aim to represent things. The concept WRONG, on this approach, is a mental representation that aims to represent the property of wrongness, whatever that turns out to be. Concepts are also presumed to be the tools by which we grasp words and the mechanisms by which words come to refer. Thus, the concept WRONG is something in my head that allows me to understand the word "wrong" and it is that in virtue of which the word refers, in fact it is not an empty expression. I will not defend this basic approach to what concepts are here (see Prinz 2002), but I want to draw out an obvious implication. If concepts are in the head, then we can study moral concepts using whatever means are available for determining what goes on psychologically when people use moral vocabulary.

This is where neuroscience and psychology come in. There are now numerous studies that are designed to discover what goes on in the head during moral reflection. Consider, for example, the barrage of recent neuroimaging studies in which researchers measure brain activity during moral reflection and compare it with brain activity during nonmoral tasks. Some studies ask subjects to deliberate on moral dilemmas, some ask them to decide whether something is morally right or wrong, some show subjects morally significant images, some have them perform morally significant actions in virtual worlds, and so on (Greene et al. 2001; Heekeren et al. 2003; King et al. 2006; Moll et al. 2002). These studies tend to show a similar pattern. In addition to brain areas associated with attribution of mental states (did the agent act intentionally?), the major players in moral cognition are brain areas associated with emotion. Common hotspots include areas such as the orbital frontal cortex and the temporal pole, which are involved in assigned emotional significance to events, and areas such as cingulate cortex, which are associated with emotional experience. This has led researchers to conclude that emotions are centrally involved in moral judgment.

Neuroscience cannot reveal which emotions are active when people make moral judgments, because we cannot yet distinguish the neural circuits underlying distinct discrete emotions. But behavioral studies can shed some light here. Experiments have been designed to determine what emotions arise in reflecting on morally significant events. The most direct way to measure this is present participants with an event and ask them what they would feel. For example, Rozin et al. (1999) gave subjects a list of moral transgressions that they had divided into three categories: crimes against persons (such as stealing or harming), crimes against

community (such as violations of status hierarchies), and crimes against nature (such as violations of sexual mores). In a forced-choice design, they found that such events elicit anger, contempt, and disgust respectively. In an unpublished follow-up study, I present subjects with a similar list of transgressions, but rather than asking how they would feel if someone else performed the immoral act, I asked how they would feel if they themselves behaved immorally. I found that crimes against persons elicit guilt, crimes against nature elicit shame, and crimes against community can elicit either. The main lesson, for present purposes, is that there are a number of different emotions that arise in the context of moral judgment. What they share in common is that they are all emotions of blame. Each feeling can be described as negative appraisal of the person of action being judged.

The studies that I have been describing so far establish that emotions arise when we make moral judgments, but they do not establish how those emotions relate to judgments. Are they mere sequelae of moral reflection, or do they play a more integral role? To address this question, experiments have been designed to show that emotions can influence moral judgments. For example, Wheatley and Haidt (2005) showed that hypnotically induced disgust amplifies judgments of moral wrongness even in response to scenarios that subjects in a control group would regard as morally benign. Similarly, Schnall et al. (forthcoming) showed that wrongness judgments sometimes increase when participants assess scenarios at a filthy desk, and compared with a clean desk.

It has also been shown that moral judgments diminish when negative emotions are reduced. Schnall et al. (submitted) showed that subjects are offer lower wrongness judgments if they are allowed to wash their hands, and Valdesolo and DeSteno (2006) have shown that watching a comedy routine dramatically reduces the judgment that it is wrong to kill one person in order to save five. Similar conclusions can be drawn from studies of psychopaths. Psychopaths are known to have a profound deficit in negative emotions such as fear and sadness, and they also have a profound deficit to understand moral judgments. Blair (1995) has evidence that psychopaths treat moral rules as if they are mere social conventions.

Putting these pieces together, we might venture that moral judgments contain an emotional component. The neuroimaging studies suggest that emotions arise when we make moral judgments. The hypnosis and dirty desk studies suggest that those emotions are not mere epiphenomena but are related to what people report. The hand washing, comedy, and psychopathy studies suggest that these emotions do not merely influence moral attitudes but may be necessary for such attitudes. Remove emotions, and people no longer judge that things are morally wrong. The evidence does not deductively entail that moral judgments contain emotions, but that conclusion explains and predicts the empirical evidence.

The studies mentioned so far provide *a posteriori* support for the conclusion that moral judgments contain emotions. By that I mean it is the kind of support that we might not discover by merely reflecting on what we mean by moral terms. Philosophers have long assumed that the structure of concepts is available to introspection. This need not be the case. For example, psychologists have used empirical evidence to support the claim that most concepts are associated with prototypes: representations of typically category instances. Evidence for this claim

includes the fact that people are faster at classifying prototypical instances and people learn to classify prototypical instances earlier in development. We might not have discovered the existence and prototypes by introspection, but that does not mean that we should deny that concepts are prototypically structured. The ubiquity of prototype effects has been used to support the hypothesis that concepts just are prototypes, or at least contain prototypes as proper parts. The argument for such a conclusion is simple: concepts are what we use to grasp words; empirical evidence suggests that we use prototypes to grasp words; therefore, concepts are prototypes. Forget what you think about the second premise and, hence, the conclusion. My point here is that empirical methods can be used to discover the structure of concepts. Thus, we can empirically discover that moral concepts have an emotional basis.

That said, the emotional basis of moral concepts might be knowable *a priori* too. We may have conscious access to this fact. To me it seems perfectly obvious that moral judgments contain emotions. The problem with this bald assertion is that philosophers tend to have different intuitions. What seems obvious to me may seem obviously false to you. This is another reason why empirical methods can be useful. When philosophers disagree about the intuitions that direct their theories, they may be victims of theoretical bias. Theories can influence our intuitions. To escape such bias, we can try to consult the intuitions of people who are not professional philosophers, such as college undergraduates.

Suppose we want to know whether people overtly recognize that emotions have an intimate link to morality. We can design a study to see whether people regard emotions as epiphenomena that are merely contingently associated with moral views, or whether instead they recognize what the empirical findings tend to support: that morals have an emotional basis. Toward this end, I asked a group of North Carolina college students to consider two scenarios (Prinz, unpublished data). One describes a person who verbally condemns marijuana smoking but has no negative emotional response to it. The other describes a person who verbally insists that marijuana smoking is morally acceptable, but feels disgusted at those who smoke and ashamed when he himself smokes. For each scenario, subjects were asked whether the person's moral values were reflected in what they said or in how they felt. In both cases, a significant and sizable majority said that the moral value is reflected in the emotion (ninety percent in the first case and seventy percent in the second case).

Recall that the studies mentioned earlier already provide empirical support for the claim that moral judgments have an emotional component. The marijuana study shows that this empirical hypothesis may even readily accessible as a conceptual truth. The marijuana study suggests that people tend to attribute moral judgments when and only when emotions are in place, and the other studies suggest that people engage in what seem to be moral judgments when and only when emotions are in place.

The marijuana study helps to adjudicate between some competing interpretations of the other results. On one interpretation, emotions exhaust or our moral concepts. The concept WRONG, on this view, is just a negative feeling (or perhaps a disposition to have a negative feeling). But some philosophers might resist a

reduction of WRONG to a mere feeling; they might suppose that moral concepts have two parts: an emotional part and some kind of descriptive part. If that were the case, it should be possible to have these two parts independently. But then, in the scenario where someone feels disgusted about marijuana smoking but insists that it is okay, subjects should not be so quick to judge that the agent is opposed to marijuana. They should take his verbal behavior as evidence for the assumption that he lacks the nonemotional component of moral judgments. Thus, subjects seem to think that emotions are sufficient for moral judgments. We can also learn something from the other scenario, in which the agent feels no negative emotions but says that marijuana smoking is wrong. When presented with the handwashing, comedy, or psychopathy studies, one might be tempted to draw a modest conclusion that emotions are not really necessary for moral judgments, but merely facilitate moral judgments. The fact that most subjects think the dispassionate marijuana critic is not really morally opposed to the drug suggests that they regard emotions as necessary to moral judgment. These observations suggest that emotions are not merely components of the concept WRONG, much less contingent components, but may actually constitute that concept in some sense. If one feels disgust and other emotions of blame toward an action, one thereby judges it to be wrong, and if one fails to have these emotions, there is nothing left to call the moral judgment. Further studies can help confirm that the constitution thesis is part of our folk theory of moral concepts.

One interesting feature of the marijuana study is that it complicates the standard philosophical distinction between *a priori* and *a posteriori* methods. When subjects fill out their responses on a questionnaire, they are using their *a priori* intuitions. But, no single individual's response provides adequate evidence for the claim that there is a conceptual link between morality and emotion. At best, we could say that the individual draws such a link. It is only when participants are averaged together that evidence for a general (that is to say, widespread) link emerges. So the link between moral judgments and emotions, though available to many though armchair reflection, is in this sense *a posteriori*. It would be helpful to introduce a third term for empirically oriented philosophy. We can say that conceptual claims made on the basis of analyzing multiple *a priori* intuitions are *a posterior priori*—made after the prior. Many experimental philosophers have conducted *a posteriori priori* studies, and in this respect their methods continue to use armchair philosophical reflection as a window into the structure of concepts. I think these methods are strongest when coupled with less direct methods for observing conceptual structure like those described above. Without studies of how we actually moralize, studies of folk intuitions about moralizing could be accused of getting morality badly wrong. Without studies of folk intuitions, studies of what we do when asked to moralized could be dismissed as showing that people are not really moralizing in those cases, but are instead doing something else (e.g., Sinnott-Armstrong [2006] argues that emotions are merely used as heuristics that stand in place of real moral deliberation).

Collectively, the studies that I have been discussing support the conclusion that moral judgments have an emotional basis, and, more specifically, that the concept WRONG is constituted by negative emotions. Elsewhere, I have elaborated

276   *Jesse J. Prinz*

on this hypothesis considerably (Prinz 2007). Here I want to focus on a possible implication.

STEP 2: FROM CONCEPTS TO REFERENCE

It is widely assumed in philosophy that if moral concepts are constituted by emotions, then they do not refer to anything. This is the kind of view that was defended by emotivists. They claimed that sentences expressing moral judgments, such as "stealing is wrong," do not aim to assert anything that can be evaluated as true or false. "Wrong" expresses an emotion, and thus "Stealing is wrong" is just an expression of a negative feeling toward stealing. Some emotivists suggest that when we say stealing is wrong we are also enjoining listeners to feel as we do about stealing, but we are not attributing any property to stealing. Emotivists seem to assume that if a word expresses an emotion it cannot also refer to a property. That assumption is confused. Emotions can refer to properties. After all, we use emotions to classify things. We can divide the world into disgusting things and nondisgusting things, scary things and nonscary things, irritating things and nonirritating things, and so on. The very fact that we use emotions to sort things suggests that emotions function as representations. Indeed, sorting can be characterized as a primitive form of reference. All things being equal, if a mental state is used to sort things into a category that can be distinguished from other categories, then that mental state refers to that category.

This is the starting place for a naturalistic theory of reference. If we think of referring a kind of activity, we can easily identify behaviors that indicate that reference is going on. If I put some items in bin A and others in bin B in a predictable way, then I am clearly distinguishing these items, and this can be characterized in semantic terms (I am representing A items and B items differently). Emotions are good classifiers. I could fill a bin with disgusting things and another bin with nondisgusting things. That suggests that disgust is in the business of representing. If I say that something is disgusting, I imply that it has the property that disgust represents.

To figure out what a mental representation represents, we need to apply our best theory of reference. Philosophers have typically developed theories of reference from the armchair. Even card-carrying naturalists about reference (those who want to characterize reference in nonsemantic terms) tend to carry on as if their project is immune to empirical evidence. The standard strategy is to assume that we have internal representations that refer to certain categories (such as natural kinds) and then see whether there could exist a relation between mind and world that uniquely maps internal states onto those categories. They try to show that certain natural relations could in principle get the semantics of our categories right. This project has its place, but I think it suffers from two limitations. First, practitioners never take the obvious next step of seeing whether the concepts we use actually do bear the specified relation to any category in the world. Suppose you think that concepts refer to what they are nomically related to (cf. Fodor 1990). It would be worthwhile to test this theory by seeing whether the concept that lights up when we use the word "dog" also bears a nomic relationship to the property of

doghood. If it does not, that would be an embarrassment to the theory. But as soon as we pose this test, we immediately get into trouble. First, it is far from clear how one can test for a law of this kind. The mental state type that is active when I use the word "dog" will probably be activated when I see some dogs and not others. Does that show that it refers to just some dogs? One should say it refers to all dogs, because the dogs that do not cause the mental state to activate belong to the same category as the dogs that do cause it to activate: they are all dogs. The semantic theorist cannot simply stipulate that the concept is nomically related to this larger category. Rather, the decision to treat the mental state as a representation of dogs seems to be a kind of decision that we make, not just as semantic theorists, but also as users of the concept. Of course, we do not have the mental resources to identify all dogs (we do not know the dog essence). But we can, as a matter of policy, treat the mental state that is activated by paradigm instances of dogs as a natural kind concept. When we do that, we are, in effect, adding another component to our theory of reference. We are saying that reference requires both nomic relations and semantic policies about how to select from the various possible categories that might be nomically related. I do not want to suggest that these policies cannot be naturalized. They probably can. I want only to say that the failure to think seriously about how to test semantic theories has led some theories to ignore the possibility that reference involves this further component.

This brings me to a second limitation of semantic theories as they are developed by naturalistic philosophers of mind. Very often, reference is treated as a purely relational property that has no bearing on the behavior or psychology. It is a fact about me, as I am situated in the world, that my concepts refer to certain things and not others, but this fact is behaviorally inert. If the concept I express by "water" referred to XYZ rather than H_2O, it would make no difference to me. By focusing on such examples, philosophers have implied that the theory of reference will have little to do with the theory of psychology (construed narrowly as a the science of behavior and processes that take place in our heads). But this is a mistake. If reference depends on semantic policies, then it depends on psychological states. Indeed, the reason why we say that my water concept refers to the local watery stuff is that I have a semantic policy to this effect. If I were totally neutral about the question of whether my water concept refers to a natural kind or to any clear liquid, then it would not refer to a natural kind. The study of reference should centrally involve an investigation of what we take our concept to refer to. Semantic policies play a role in determining reference, and such policies lead themselves to empirical investigation. We can ask people, "What kind of things are dogs?" or "Could something be a dog if it didn't look like one?" Again, one can think of referring as a kind of activity. It is a complex activity that involves both sorting things and forming beliefs about the nature of the things we sort. Theories of reference should be tested and they should not be seen as psychologically inert.

In semantic theory, the term "naturalism" refers to attempts to provide reductive explanations. In epistemology and moral psychology, it refers to attempts to ground philosophical theories in empirical evidence. What I am suggesting is that semantic theory should be naturalized in this second sense. A theory of reference should be seen as a psychological theory—a theory of mental processes

and behaviors. What our concepts refer to is up to us, and if we want to figure out what reference is, we should see what we say about reference and how we engage in the activity of referring.

Taking these lessons on board, we can see how one would go about to determine what, if anything, our moral concepts refer to. First, we need to see what things we sort under our moral concepts. If, as I have suggested, moral concepts have an emotional basis, this amounts to the question of what things induce the relevant emotions (anger, contempt, guilt, etc.). Second, we need to see what are semantic policies that are regarding moral concepts in order to see how to extend and delimit the class of things that elicit these emotions.

There has been a considerable amount of empirical research on what elicits each of our moral emotions (see Haidt 2003, for a review). One set of findings was already mentioned above: anger and guilt are elicits by crimes against persons; disgust and shame are elicited by crimes against nature; and contempt, guilt, and shame are elicited by crimes against community. Many other studies add details and refinements to this big picture. Guilt is most readily induced when the person harmed is someone we care about (Baumeister et al. 1994). Disgust can be induced by crimes against persons when those crimes are especially senseless, or violent, or motivated by demeaning attitudes such as racism (Haidt et al. 1997). Shame often arises when someone imagines being judged by an audience (e.g., Ferguson et al. 1991).

As these observations already make clear, the class of things we call "morally wrong" is something of a hodgepodge. It includes such varia as theft, rudeness, and bestiality. In this respect, WRONG is quite unlike DOG, SPOON, or POTATO. These concepts refer to categories whose members vary, to some extent, but it is fairly easy to identify intracategorial commonalities. Moreover, there is considerable cross-cultural variation in what elicits our moral emotions and gets labeled morally wrong. We morally condemn infanticide, cannibalism, slavery, honor killing, torture, polygamy, first-cousin marriage, and male dominance, but these practices are or have been widely accepted by other cultures. Even within our society, there is considerable variation: Southerners are more likely to favor corporal punishment (Nisbett and Cohen 1996), liberals and conservatives differ on whether distribution should aim for equality or equity, and members of various religious groups diverge on questions about abortion and birth control. By comparison, there is little cultural disagreement, I suspect, about what counts as a potato.

Moral variation poses a challenge for anyone aiming to explain what our moral concepts refer to. With natural kind concepts and many artifact concepts, a large portion of the things that we classify under a single label share some unifying properties. We can discover what category members have in common. It sometimes takes years of scientific research to do so, and there is often enough variation in the category that scientists are forced, at certain junctures, to simply stipulate how a term should be used (are whales fish?). But it is, nevertheless, fairly easy to see how unifying features might be identified, and there is plenty of reason for optimism about unifying projects before the search for essences begins. Members of natural kind concepts typically *seem* similar on the face of it, and there is considerable consensus about what objects fall into what categories.

When we encounter a category that is extremely heterogeneous on the face of it, there are four possibilities. One possibility is that there is a unifying essence that just is not obvious, but could be discovered through hard work. In the case or morality, that seems highly unlikely. It is not even clear how we would go about finding a unifying essence. Should we put stealing and bestiality under a microscope and see what they share in common? If different people classify different things as wrong, whose class of wrongs should we consider in our hunt for essences? One might think that the search for the essence of wrongness has been a central task in moral philosophy. Perhaps normative ethicists are like the chemists of the deontic domain. I am skeptical of this interpretation. I think normative ethicists are rarely looking for an essence that would unify moral judgments. Rather, they tend to abandon many ordinary intuitions about what is wrong and build general theories based on a handful of cases. In the West, moral philosophers have typically focused on harm norms, ignoring the fact that ordinary people often condemn acts in which no one is harmed (such as consensual incest). Some moral philosophers focus on moral rules that involve character, deliberately ignoring or postponing consideration of rules that pertain to action. Others do the opposite. The search for common essence seems as fruitless as using a washcloth for a blanket: pull it up to cover one part of your body, and the rest of your body will be exposed. If normative ethicists are trying to unify the class of things we moralize, they are not succeeding. More often, I think they are trying to develop new systems of values: abandon your folk moral attitudes, they say, and follow a more coherent set of action guiding principles. That is a noble project, but it will not help with the question at hand: What do our ordinary moral concepts refer to? My present point is that ordinary moral concepts probably do not refer to a unifying essence.

This brings us to a second possibility. Perhaps moral concepts simply fail to refer. If there is no intrinsic essence unifying the majority of things we classify as WRONG, then maybe that concept is empty. Suppose I introduce the word "proggle" to refer to the category containing potatoes, spoons, and dogs. Suppose that you pick up the term but use it for potatoes, spoons, and spiders. Someone else uses "proggle" when talking about eggplants, spoons, and gorillas. At a certain point one might ask what proggle refers to. There is certainly no intrinsic essence, even as any one individual uses the word. That might raise the specter of elimination: there are no proggles. Whether this consequence follows depends the semantic policies underlying the use of the PROGGLE concept. Suppose I intend the concept to refer to a natural kind. If so, the fact that there is no intrinsic essence in my category suggests that my concept is incoherent. My classificatory behaviors cannot be reconciled with my semantic policies. From this it would follow that my concept does not refer.

Can a similar conclusion be drawn in the case of moral concepts? That depends on our semantic policies toward moral terms. Figuring out what those policies are is an empirical project. We need to investigate whether moral concepts are treated as natural kind concepts or in some other way. Do people assume that there is a unifying essence to the moral? If so, is this just a dispensable belief or is it integral to how people use moral concepts? Undoubtedly some people think there is an essence to morality. For example, some think that all moral rules issue

280   *Jesse J. Prinz*

from divine degree. But three things are noteworthy about that particular belief. First, it is not universal. Second, those who have it would concede that something could still count as morally wrong if God had not commanded against it (Nucci 2001). Third, divine decree is not an intrinsic essence, so those who take this to be the unifying feature of moral rules do not thereby assume that morals constitute a natural kind. Indeed, there is good reason to suppose that people do not ordinarily think of wrongness the way they think of tigers or gold. Unlike those categories, we do not establish sciences for discovering the essence of morality (this article not withstanding), and many of us do not feel compelled to consult an expert when deciding whether something is really good. With natural kinds, every alleged instance of the category can potentially be questioned: Are Chihuahuas *really* dogs? With morals, some cases seem to be non-negotiable, though these will differ from person to person. If I read a headline reporting that experts had discovered that rape is not really wrong, I would assume it was a very sick joke. Basic values just do not seem contestable. And when they are contested, we do not respond by calling in experts. Rather we reflect deeply to make sure that we really have the values that we think we have. Is my moral abhorrence of rape really basic? And if it is not, might it be up for debate after all? I sincerely doubt the answer is yes, but to find out, I need to look inward into my values, not merely outward at the world. In sum, I think we should empirically investigate whether people treat moral concepts as natural kind concepts, and I suspect the answer will be negative.

If moral concepts are not natural kind concepts, then variation within the range of things that people moralize will not necessarily entail conceptual inco-herence and vacuity. A third possibility (beyond intrinsic essences and vacuity) is that WRONG is a disjunctive concept, revering to some fixed list of basic wrong, and other wrong derived from those. Many traditions have lists of basic wrongs (the Ten Commandments), and we certainly possess some disjunctive concepts (recall the philosophical literature on JADE). What troubles me about this proposal is that disjunctive concepts are rare, and they usually arise in cases where we were aiming for a natural kind and ended up pointing to objects that lack intrinsic unity. JADE is a disjunctive concept because people thought that nephrite and jadeite were intrin-sically alike, and that assumption proved false. When that was discovered, we probably should have become eliminativists about jade (the concept turned out to be incoherent), but it proved useful to drop the natural kind policy, and preserve a concept that referred to two stones rather than one. In the case or moral categories, the disunity is quite apparent, unlike jade. It hardly seems that we made a mistake and thought that bestiality and promise breaking were intrinsically alike. More-over, we need a story about why we co-classify such heterogeneous behaviors in the first place. We have such a story with JADE. If we lack such a story with WRONG, it would seem that we arbitrarily grouped together a motley disjunction and smacked on a single label (recall the proggle case). But if there is some feature in virtue of which we co-classify wrongs, then perhaps we should say that feature, whatever it is, provides the sought after categorical unity. The point can be made in terms of semantic policies: absent any evidence that we simply think or moral wrongs as an arbitrary list, we should look for an alternate account of how we think about this category.

If bestiality and promise breaking do not share intrinsic properties, how is it that we co-classify them. This question has both a proximal answer, a distal answer, and a really distal answer. The proximal answer is already anticipated by the earlier discussion. These wrongs, and all others, are united by the fact that they cause a certain class of emotions in us, emotions of blame. We would feel an other-blaming emotion if we caught someone sleeping with a sheep or breaking a promise, and we would feel a self-blaming emotion if we were caught performing these act ourselves. In this respect, very different wrongs feel alike. By comparison, consider the range of things we find funny or the range of foods we find delicious. A subtle pun and a slapstick routine may have no interesting intrinsic properties in common, but they both amuse us. A serving of tiramisu tastes nothing like the squid ink pasta in the first course, but both cause gustatory pleasure. Very different things can elicit similar feelings, and when that happens, we group those different things together. In the moral case, we can ask why such different things cause such similar emotions. This brings us to the distal story. Different actions elicit similar feelings because, during the course of moral education, we get punished for very different things. Punishment takes a number of different forms, from physical discipline to social ostracism, but all these different forms have a common effect: they condition emotional responses. When children are corrected for misdeeds, the methods of correction are emotionally evocative. Caregivers also instill emotions by expressing outrage when morally questionable behavior is discussed. If the topic of bestiality come up, you can be sure that strong emotions will be expressed. Moral learners pick up on these emotions and acquire emotional dispositions. Why do caregivers condemn some actions and not others? The short answer is that they were emotionally conditioned to condemn those things, and the long answer (the really distal story) is that every value in a given culture has a genealogy. Many were introduced to promote social cohesion, some came about through happenstance, others reflect bygone superstitions, and so on. The details do not matter here (see Prinz 2007, for some examples). What matters is that there is always an explanation for why something gets moralized, and moralization consists in the cultivation of emotional dispositions. When we classify two different things as wrong, we do so on the basis of the emotions they elicit.

This suggests a fourth proposal about the semantic consequences of moral diversity. We could say that WRONG refers to a response-dependent property: the property of causing emotions of self-blame and other blame. Other philosophers have made similar proposals, most notably John McDowell (1985) and David Wiggins (1987). They liken wrongness to funniness: the funny is, arguably, that which causes feelings of amusement. In order to defend a response-dependent account of what moral concepts designate, it would be good to show that this conclusion follows from an independently plausible account of how concepts refer. I suggested that reference often involves two factors: a pattern of nomic co-occurrence and a semantic policy. It is trivially true that the property of causing emotions of blame causes emotions of blame in a law-like way. The property can be defined in that way. But in order for moral concepts to refer to this response-dependent property, we also have to have a semantic policy that has that implication. This is harder to establish.

282   *Jesse J. Prinz*

I have argued that we do not treat moral concepts as natural kind concepts or as disjunctive concepts. But is it really plausible that people think the wrong is just that which causes feelings of disapproval? I think it is more likely that we have a flexible semantic policy when it comes to moral concept. If it turned out that there was a mind-external source of morality (such as a divine decree), we would happily say that morality is an objective feature of the world. But if there is no objective feature, we are happy to treat moral facts as subjective. Response-dependence is a kind of fallback plan. One could garner empirical support for this claim by asking people (in jargon-free language) whether failure to find a mind-independent source of morality would lead them to conclude that moral claims are vacuous. I do not know for sure how such a study would come out in the moral case, but I take the outcome as quite predictable in the case of concepts such as FUNNY, DELICIOUS, or BEAUTIFUL. Suppose there is no intrinsic essence to things that amuse us; should we say that nothing is funny? Suppose there is no essence to things we find delicious or beautiful; should we say these properties do not exist? No. Instead we people have different senses of humor, we say that food preferences are a matter of taste, and beauty is in the eye of the beholder. These platitudes reveal a tendency to treat emotionally based concepts as response-dependent, when we find that they do not correspond to mind-independent features of the world. The same fallback strategy may be in place with color concepts. Locke argued that these refer to secondary qualities—powers to cause certain experiences in us.

There is a plausible diagnosis of why we seem to default to response-dependence in all these cases. Even if we think surfaces are really colored and good jokes are intrinsically funny, we also know that color and funniness are manifest features of the world. They are not hidden posits that we can reveal only through scientific instruments. In this they differ markedly from concepts such as WITCH, FLOGISTON, and other concepts introduced to refer to (especially unobservable) natural kinds. We may have a semantic policy that says something like "save appearances!" If some apparent property turns out not to have a mind-independent essence, we do not say the apparent property does not exist. Instead, we say it is merely apparent; in other words, response-dependent. This explains why people continue to believe in the existence of colors, beauty, and funniness.

So the trick to determine whether moral properties are response-dependent is not to see whether people antecedently believe that. Some people may think morality has a mind-independent source. The real question is whether people think of morality as a manifest feature of the world. Do they think that any empirical discovery could show that morality does not exist? Would the discovery that moral values come from us entail that there are no moral truths? To my knowledge, no one has directly investigated lay opinions about such questions. One piece of evidence comes from the fact that there seems to be considerable diversity of opinion about where morality comes from. Some claim to believe in divine command, others do not. Some people are moral absolutists, and some are relativists. In a study of moral objectivist, Nichols (2004) showed that people are divided on the question of whether there is a fact of the matter who is right in moral disputes. Kelly et al. (2007) found similar results, with over half of their subjects ⑦ saying that it is okay to hit someone if the local authorities allow it. The individual

differences suggest that moral concepts are semantically flexible. Some people have strong convictions about whether morality is mind-independent or mind-dependent property, but these convictions are not semantic policies. They are just collateral beliefs about the nature of moral properties. If someone is wrong about whether morality depends on human responses, that does not entail eliminativism. In other words, the variation is that beliefs about the nature of morality can be taken as evidence for the view that we regard morality as a manifest kind. And, if that is right, we will try to save appearances if we discovery that there are no mind-independent moral facts.

More evidence is needed before any confident conclusions can be drawn about moral reference, but let me sum up with the conclusions that seem most plausible to me. First, the diversity of moral values problematizes efforts to identify a mind-independent essence of moral wrongs. So I do don't think WRONG refers to a mind-independent property. Second, that does not entail eliminativism, because the assumption of mind-independence is not a semantic policy when it comes to moral concepts. Third, the best account of what WRONG refers to, given our semantic policies and the range of things we call "wrong," is a response-dependent account: WRONG refers to the property of causing emotions of self-blame and other-blame. Such an account would need to be refined in various ways (whose response matter? under what conditions?), but it will serve for present purposes (see Prinz 2007).

STEP 3: FROM MORAL TRUTH TO OBLIGATIONS

Suppose I am right, and WRONG refers to a response-dependent property. It then follows that there are two ways to figure out what is wrong. First, there is a first-person method. If something seems to cause emotions of blame in me that is direct evidence that it is wrong. It is not decisive evidence. I may misinterpret my feelings, or those feelings could derive from an extraneous source, such as a dirty desk. But, all things being equal, if something seems to make me feel angry and guilty, it probably has the disposition to cause these emotions in me, and that entails that it is wrong. From a first-person perspective, we use our emotions as a guide to morality, and they are a pretty reliable guide because moral properties are constituted by powers to cause emotions. Phenomenologically, we do not experience this as an inference. We do not typically say to ourselves, "that made me mad, so it must be wrong." Rather, the anger constitutes our belief that something is wrong. So, from a first-person perspective, there is often an immediate, causal transition from perception of an event to representation of that event as wrong, by means of an emotional response. Thus, moral judgments appear to us as intuitions, rather than inferences. Put another way, the state transition from perceiving an event to condemning it is not a deductively valid inference; it is a causal chain. So, from the first-person perspective, we do not *deduce* moral facts from descriptive facts.

But there is also a second way to figure out whether something is wrong. From a third-person perspective, we can investigate what things in the world are disposed to cause emotions of blame. One could adopt this third-person

284   *Jesse J. Prinz*

perspective with regard to oneself. For example, a therapist might help me see that I get mad when my spouse stays out late at night, and this insight might lead me to conclude that I must be regarding that behavior as wrong. If I also believe that WRONG refers to that which is disposed to cause emotions of blame, I may conclude that my spouse really has done something wrong.

This kind of inference can also take place when we are investigating other people. Anthropologists studying the Gisu of Uganda have discovered that members of this group show moral outrage when someone does not avoid his or her mother-in-law (Heald 1990). If wrongness is the property of being disposed to cause such responses, then the anthropologist can conclude that it is wrong for members of the Gisu to visit their spouses' mothers.

When we make inference of this kind, we are discovering wrongs empirically. Purely descriptive premises go into the discovery process, and these deductively entail that something is wrong. We observe that someone has an emotional response to something; we know that wrongness consists in causing such a response; and we infer that the thing causing the response is wrong. The argument makes use of the semantic theory I argued for in the last section, but a similar argument could be constructed for different theories of reference. For any theory of reference and for any theory of concepts, there is some relation R that obtains between a concept and what it represents. If it is a naturalistic theory (in either of the senses mentioned above), one can empirically investigate what bears R to the concept WRONG. Thus, one can empirically discover wrongs.

It follows trivially that one can empirically discover obligations. I do not want to offer a complete analysis of the concept OUGHT here, but one entailment relation strikes me as relatively uncontroversial: if something is wrong, then one ought not do it. If one can discover wrongs empirically, and one can derive oughts from wrongs, then one can discover oughts empirically. One can derive an ought from an is.

The argument that I have been developing is easy to summarize. It goes like this:

P1. The concept WRONG is constituted by emotions.
P2. Emotions refer to response-dependent properties.
P3. There is some person P, such that some type of action A, disposes P to feel the emotion constituting the concept WRONG.
C1. Therefore, doing A is wrong for P.
P4. "It is wrong for *x* to do *F*" entails "*x* ought to refrain from doing *F*."
C2. Therefore, P ought to refrain from doing A.

I suspect many readers will not be convinced by this argument. I have given only preliminary arguments for the core premises, but, as I just noted, a similar argument could be constructed for different theories of moral concepts and moral semantics. For that reason, I will not spend time reviewing objections that focus too heavily on specific premises. But I must consider two objections. According to the first, the conclusion of this argument may seem so obviously wrong that no argument could possibly support it. According to the second, the conclusion can only be defending on a reading that robs it of normative force.

OBJECTION 1: OBLIGATING EVIL

The first objection I want to consider concerns the relativist implications of the view that I have been defending. If the truth of wrongness judgments depends on emotional dispositions, and different people are disposed to have different emotional responses, then the truth moral judgments will vary from individual to individual. I think this is an unavoidable consequence of the theory, and one that we must be prepared to accept if we are open to the idea that the semantics of moral concepts can be empirically studied. Here, as elsewhere, we are hostage to empirical fortune, and I think relativism is supported by the evidence. Relativism is a hugely controversial thesis, which I cannot defend here (see Prinz 2007). But, there is a worry that I must address here. On the face of it, the relativism entailed by my theory of moral concepts seems to undercut my argument for deriving obligations from descriptive premises. That would be a serious embarrassment.

Here is the worry. We can empirically discover that members of some other group take themselves to be obligated to do something. For example, we can discover that the Gisu take themselves to be obligated to avoid their mother-in-laws. But from this it does not follow that they ought to avoid their mother-in-laws. The point becomes all the more obvious when we consider cultural groups that have values that we find abhorrent. Can we say of the extremist Rwandan Hutus that they ought to have massacred the Tutsis? Can we say of those who engage in honor killings, such as some traditional members of the Yedizi religious group, that they ought to stone their sister's and daughter's to death when they become romantically involved with members of other groups? Can the mere fact that someone sincerely believes that they are obligated to do something horrific entail that the person really ought to? If not, then something has gone wrong with the argument above.

I think this worry is more moral than philosophical. We do not like the idea that people are obligated to do things that we find morally repellent. If the theory has that consequence, we will regard it as very unfortunate, but that does not make the theory wrong. To make the objection stick, one would have to convert disdain for the conclusion into a substantive objection. One strategy for doing that is to argue that the conclusion of the argument can generate contradictions. The argument suggests that people ought to engage in honor killings, after all failure to engage in that practice causes strong emotions of blame in people who abide by that custom. But it also follows from the argument that one ought not engage in honor killing. Some people (myself included) find the practice reprehensible, and that implies that it is wrong, and if it is wrong, one ought not do it. This seems like a contradictory pair of entailments.

This objection is not decisive, because the apparent contradiction can be eliminated in three ways. First, one can follow Gilbert Harman (1975) and say that the word "ought" is semantically restricted so as to apply only when the person to whom it is applied possesses the values in question. On this interpretation, we cannot say that the Yedizi ought to refrain from honor killings, so we cannot generate a contradiction. Second, one can relativize obligations to value systems. We can say that the Yedizi are obligated to engage in honor killings on their value

286   *Jesse J. Prinz*

system and not ours. Third, we can eliminate the contradiction by placing the negation inside the obligation. In other words, we can say that the Yedizi are obligated to engage in honor killing and not to engage in honor killing. This formulation escapes the *logical* contradiction, even if it places the Yedizi under *practically* inconsistent requirements. The narrow scope of "not" is justified on the grounds that saying "not obligated" implies that there is no obligation, which is, if the semantic theory is right, false.

But there is another semantic worry that is harder to avoid. Arguable, the concept OUGHT can only be used to express one's own values. Like WRONG, the concept OUGHT may have an emotional basis. When we say that someone ought to do something, we may be expressing how we feel. I certainly do not feel good about honor killing, and, consequently, it is not just odd for me to say that people ought to engage in this practice; such a statement would actually semantically anomalous. It would be equivalent to me saying that fine Bordeaux wines are yucky. A child might call a Bordeaux "yucky" but I cannot apply this word without feeling the disgust that the word conventionally expresses. Likewise, I cannot say someone ought to engage in honor killing if I do not have the feelings that conventionally constitute ought thoughts. (And, perhaps I cannot say that something is wrong unless it is wrong according to my values.) If this intuition about the concept of OUGHT is right, then there cannot be any deductive argument that uses "ought" as the main verb in a conclusion without having some other evaluative term used in the premises. There can be no derivation of an ought from an is.

This is a powerful objection, but it can be answered. I offer three lines of reply, which range from very concessive to not concessive at all. First, I might concede that no ought can be derived from an is, while maintaining something weaker: we can derive an *obligation* from an is. Perhaps we can say that people who support honor killings are under an obligation to kill, while denying that they ought to kill. To my ear, this does not sound inconsistent. On this reply, I can continue to claim that obligations are empirically discoverable, which is an interesting result.

The second reply is slightly less concessive. I might concede that we cannot derive conclusions about what other people ought to do, but I could derive conclusions about what I ought to do. Recall that we can take a third-person stance toward ourselves. Suppose that I am undergoing a severe depression, and as a result I am not feeling strong emotions of any kind, whether negative or positive (see Mele 1996). I can still discover, say by reflecting on my past behavior, that certain actions are disposed, when I am healthy, to cause emotions of blame. Knowing this fact about myself, I can infer that those actions are wrong, and that I ought to avoid them. Here, the "ought" is semantically licensed because I am disposed, under better conditions, to feel the emotions that normally accompany the word. If this reply is right, the one can derive an ought from an is, as long as the argument is stated in the first-person.

The third reply is least concessive. Perhaps there are two uses of "ought," one of which is expressive, the other of which is not. By analogy, there may be two uses of delicious. I can say that chicha (a fermented beverage made from corn and saliva) is delicious to Amazonian people while denying that it is delicious to me. This use of the word does not express any feelings. But if I drink some Bordeaux

as say "that's delicious!" I have thereby expressed how I feel. Notice that this expressive use would be the default reading, for pragmatic reasons, if I did not specify a context. If I say, "Chichi is delicious," I imply that it is delicious to me. But I can cancel this implication by saying that it is delicious to Amazonian people. Likewise, there may be a nonexpressive use of ought. I may be able to cancel the implicature that I am deploying the expressive use by saying, "You ought to avoid your mother-in-law if you are a member of the Gisu." If "ought" has a nonexpressive use, then I can derive an ought from an is. It is an empirical question whether the word "ought" ever allows such a nonexpressive use. One might investigate by polling intuitions about sentences like the one I just formulated about the Gisu.

I am confident about the first two replies, and less sure about the third. At a minimum, then, I would say we can derive an ought from an is when talking about our own values, and we can derive an obligation from an is when talking about people whose values differ from our own.

OBLIGATION 2: ANEMIC OUGHTS

I have been arguing that we can use empirical methods to discover obligations. Purely descriptive premises reveal the demands that moral values place on us. Critics may find my derivation of norms somewhat unsatisfying. In my introduction, I hinted that empirical findings could be used to draw normative conclusions. A normative conclusion is a claim about what we ought to do. The argument that I have put forward seems both too powerful and too weak for that purpose. It is too powerful because it seems to prove too much. It can be used to show that all of us are obligated to do whatever anyone thinks we should do. If someone thinks that I am morally obligated to indulge in cannibal feasts every February, then the argument that I put forward seems to entail that there is such an obligation. But it would make a joke of normative theory to say that we are all obligated to do everything that any moral system required. In this sense, the argument is too strong to do any work in normative ethics. On the other hand, it is also too weak. I conceded in the last section that the argument I have put forward cannot be used to adduce an other-directed ought judgment stated with emotional force. This is the kernel of truth in Hume's law: you cannot derive an oomph from an is. If the oughts that one can derive have no oomph, then they cannot be taken as recommendations for action. When I say that the Gisu ought to avoid the mother-in-laws, I am not recommending that they do that. I am, in this sense, not really making a normative claim. So perhaps my introduction was an exercise in false advertising. Perhaps I have not shown how to draw normative conclusions after all. The oughts that can be deduced from empirical premises are too anemic.

This objection reveals a genuine limitation of the argument that I have been defended, but there is also an available reply. I just implied that my argument for deriving oughts entails that everyone is obligated to do everything that anyone happens to think they should do. But this is actually a mistake. As formulated above, the argument is actually restricted to norms that are possessed by the person upon whom the derived obligation is placed. The argument entails that if a person,

288   *Jesse J. Prinz*

P, is disposed to have a certain emotional response to an action, A, then P ought not do A. It is silent on the question of whether P falls under obligations that she or he would not endorse. I have not said that every agent ought to do whatever is required on other agents' moral theories. The argument does not require that I ought to avoid my mother-in-law, for example. It entails only that each person ought to do what his or her own values require.

This restriction provides a response to the normativity objection. Notice, that from P's point of view, P ought to do what P's norms require. So the ought in the conclusion has normative force in one important sense: it is an ought that would be psychologically compelling to P. Trivially, we feel obligated to follow our own values. Indeed, we are bound by our values even when we do not realize what our values require of us. If P does not realize she ought to perform some action by her own system of values, it still follows that she ought to perform that action; she is committed to norms that carry that requirement. Even if the "ought" in the conclusion is not, semantically speaking, a normative ought (because the person drawing the inference may not have the requite feelings), it is an ought that has normative jurisdiction over the person to whom it applies. So whenever the argument is applied there is a corresponding ought with oomph.

This kind of normativity is weaker than the kind that many philosophers seek. Philosophers often want norms that are binding *regardless* of our desires and inclinations. I am skeptical that there are such norms in the moral domain. My skepticism stems from my conviction that moral concepts refer to response-dependent properties. If the arguments for that claim succeed, then the idea of moral norms that transcend our inclinations will be difficult to sustain. Thus, I think the revised formulation of the argument may yield conclusions that track the only kind of normativity we should expect to find in the moral domain. The conclusion of the argument might not be stated in an emotionally expressive way by the person making the argument, but the ought would be binding.

This suggestion may seem disturbing at first. I have suggested that there is a kind of normativity such that everyone is bound by the values they possess. The Gisu are required not to visit their mother-in-laws, and members of the Yedizi are required to kill those who dishonor them. That is an unsettling consequence of a theory. But notice that when I observe the existence of these requirements, I am not expressing my own values. I think honor killing is a ghastly institution. Indeed, I have been arguing that all of us are bound by our own values, so I am bound to condemn honor killing.

Against this suggestion, one might object that people can question whether they are really bound by the values that they happen to endorse. A member of the Yedizi can say, "I know I morally value honor killing, but should I? Is it really something I am obligated to do?" It is a consequence of my theory that this kind of question is really incoherent. That may look like a fatal objection to the theory, but it is not an embarrassment if there is a reasonable explanation of why such questions seem intelligible to the people who formulate them. Fortunately, a reasonable explanation is available. Recall that I am not committed to the view that people realize morality is a response-dependent property. People can (and often do) believe that morality is mind-independent. Someone who endorses a divine

command theory, for example, can wonder, "Are the values I endorse really God's requirements or were they planted by some kind of devil?" Someone who wonders whether her or his values are binding may be in the grip of an erroneous theory of where values come from.

There is also another explanation of how we can question of values. Suppose I come to recognize that my values are response-dependent. After I make this discovery, I can still wonder whether the values I have are the best possible values that I could have. For example, I could wonder whether my moral values lead me to flourish. I can then ask, in a prudential voice: "Should I do what my morals tell me to do, or should I pursue actions that will bring me greater well-being?" I may also recognize that I have inconsistent values, and that will lead me to think that I am morally obligated to refrain from doing something that I am also obligated to do. A member of the Yedizi might believe in the basic sanctity of human life, and this could lead to the thought: "I morally value honor killing, but I ought not engage in that practice because I am also morally obligated to protect human life." In sum, I think we are morally obligated (by our own standards!) to do whatever our moral values require of us, but we sometimes fall under competing obligations because we are morally inconsistent or because we possess nonmoral values that are difficult to reconcile with morality. Thus, the fact that people question whether they should do what their moral values demand is easily explained. It can reflect an understandable mistake about moral ontology, or it can reflect the recognition that values compete.

The process of reviewing our values, searching for inconsistencies, and deciding what to do is a central part of normative ethics. The crucial thing to notice about this process is that we do it from within a system of values. We typically begin with something that we already value and try to draw out its implications. We cannot adopt a transcendental stance, and if we could, we would not feel compelled to endorse the values that emerged from such a stance. I think philosophical work in normative ethics is an internal, value-expressive project, and normative theories are usually systematic extrapolations of values that the intended readers of such theories already endorse.

When building normative theories, philosophers sometimes select a small set of antecedently accepted values, and those values are usually discovered through introspection. Introspection is essentially an empirical method. We access our mental states through a process of internal observation. But introspection is a limited method; we do not always observe accurately, we are prone to theory-laden interpretations; we do not have conscious access to all of our thoughts, and the values revealed by introspection may be unique to the introspector. Moreover, the practice of cherry picking from antecedently accepted values results in normative theories that are parochial: they capture some of our moral commitments at the expense of others.

I think philosophical approaches to normative theory construction can be supplemented by the methods used in the empirical sciences. Experimental psychology, for example, can help us discover our normative commitments. Psychologists do this in a variety of ways: sometimes they ask people to rate morally significant scenarios, sometimes they measure emotional responses, sometimes

290   *Jesse J. Prinz*

they observe behavior, and so on. We can discover some of our own values in this way, and in some cases we are surprised. We may find that our values are inconsistent; we may find that some values are more fundamental than others; we may discover the comparative weightings of values; and we may identify factors (such as dirty desks) that skew our value judgments. All of these discoveries would be relevant to constructing normative theories. Normative ethics can systematize the values we already possess, and reduce inconsistencies by determining which of two competing values is more important to us. I think we should actively pursue all methods of discovering what our values are.

Put differently, the argument for deriving oughts takes on a special role in the first-person case. When we discover what we value, we discover the norms that bind us—the things that, by our own standards, we ought to do. In other words, when we empirically investigate the things that arouse our emotions of blame, we are contributing to a normative project. The argument shows that we can discover normative oughts. When the agent in question is me, the argument shows that I can discover what I ought to do on the basis of empirically discovered facts. The adage that descriptive methods cannot contribute to normative projects is a mistake. The psychological sciences reveal not only what we do value but also, thereby, what we ought to do. The idea that there are moral oughts that exist independently of what we do in fact care about may be a myth. If so, we may benefit from reconceiving normative ethics as a kind of auto-anthropology.

REFERENCES

Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. 1994. "Guilt: An Interpersonal Approach." *Psychological Bulletin* 115: 243–67.

Blair, R. J. R. 1995. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57: 1–29.

Ferguson, T. J., Stegge, H., and Damhuis, I. 1991. "Children's Understanding of Guilt and Shame." *Child Development* 62: 827–39.

Fodor, J. A. 1990. *A theory of content*. Cambridge, MA: MIT Press.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293: 2105–8.

Haidt, J. 2003. "The Moral Emotions." In *Handbook of affective sciences*, eds. R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, 852–70. Oxford: Oxford University Press.

Haidt, J., Rozin, P., McCauley, C., and Imada, S. 1997. "Body, Psyche, and Culture: The Relationship of Disgust to Morality." *Psychology and Developing Societies* 9: 107–31.

Harman, G. 1975. "Moral Relativism Defended." *Philosophical Review* 84: 3–22.

Heald, S. 1990. "Joking and Avoidance, Hostility and Incest: An Essay on Gisu Moral Categories." *Man* 25: 377–92.

Kelly, D., Stich, S., Haley, K., Eng, S., and Fessler, D. 2007. "Harm, Affect and the Moral/Conventional Distinction." *Mind and Language* 22: 117–31.

King, J. A., Blair, R. J. R., Mitchell, D. G. V., Dolan, R. J., and Burgess, N. 2006. "Doing the Right Thing: A Common Neural Circuit for Appropriate Violent or Compassionate Behavior." *Neuroimage* 30: 1069–76.

McDowell, J. 1985. "Values and Secondary Qualities." In *Morality and objectivity*, ed. T. Honderich, ••–••. London: Routledge & Kegan Paul.

Mele, A. R. 1996. "Internalist Moral Cognitivism and Listlessness." *Ethics* 106: 727–53.

Moll, J., de Oliviera-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., and Pessoa, L. 2002. "The Neural Correlates of Moral Sensitivity: A Functional Magnetic

Resonance Imaging Investigation of Basic and Moral Emotions." *Journal of Neuroscience* 22: 2730–36.

Nichols, S. 2004. "After Objectivity: An Empirical Study of Moral Judgment." *Philosophical Psychology* 17: 5–28.

Nisbett, R. E., and Cohen, D. 1996. *Culture of Honor: The Psychology of Violence in the South.* Boulder, CO: Westview Press.

Nucci, L. P. 2001. *Education in the Moral Domain*. Cambridge: Cambridge University Press.

Prinz, J. J. 2002. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

———. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Rozin, P., Lowery, L., Imada, S., and Haidt, J. 1999. "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)." *Journal of Personality and Social Psychology* 76: 574–86.

Schnall, S., Haidt, J., and Clore, G. Submitted. "Irrelevant Disgust Makes Moral Judgment More Severe, for Those Who Listen to Their Bodies." ⑨

Sinnott-Armstrong, W. 2006. "Moral Intuitionism Meets Empirical Psychology." In *Metaethics after Moore*, eds. T. Horgan and M. Timmons, 339–65. New York: Oxford University Press.

Valdesolo, P., and DeSteno, D. 2006. "Manipulations of Emotional Context Shape Moral Judgment." *Psychological Science* 17: 476–77.

Wheatley, T., and Haidt, J. 2005. "Hypnotically Induced Disgust Makes Moral Judgments More Severe." *Psychological Science* 16: 780–84.

Wiggins D. 1987. "A Sensible Subjectivism." In *Needs Values, Truth: Essays in the Philosophy of Value*, ed. ••, 185–214. Oxford: Blackwell. ⑩