Here are two of the biggest questions in moral psychology: (1) Where do moral beliefs and motivations come from? (2) How does moral judgment work? All other questions are easy, or at least easier, once you have clear answers to these two questions.

Here are our answers: (1) Moral beliefs and motivations come from a small set of intuitions that evolution has prepared the human mind to develop; these intuitions then enable and constrain the social construction of virtues and values, and (2) moral judgment is a product of quick and automatic intuitions that then give rise to slow, conscious moral reasoning. Our approach is therefore some kind of intuitionism. However, there is more: moral reasoning done by an individual is usually devoted to finding reasons to support the individual's intuitions, but moral reasons passed between people have a causal force. Moral discussion is a kind of distributed reasoning, and moral claims and justifications have important effects on individuals and societies. We believe that moral judgment is best understood as a social process, not as a private act of cognition. We therefore call our model the social intuitionist model (SIM). Please don't forget the social part of the model, or you will think that we think that morality is just blind instinct, no smarter than lust. You will accuse us of denying any causal role for moral reasoning or for culture, and you will feel that our theory is a threat to human dignity, to the possibility of moral change, or to the notion that philosophers have any useful role to play in our moral lives (see the debate between Saltzstein & Kasachkoff, 2004, vs. Haidt, 2004). Unfortunately, if our theory is correct, once you get angry at us, we will no longer be able to persuade you with the many good reasons we are planning on giving you below. So please, don't forget the social part.

In the pages that follow we will try to answer six questions. We begin with the big two, for which our answer is the SIM. We follow up with Question 3: What is the evidence for the SIM? We then address three questions
that we believe become answerable in a coherent and consistent way via the SIM. Question 4: What exactly are the moral intuitions? Question 5: How does morality develop? And Question 6: Why do people vary in their morality? Next we get cautious and consider some limitations of the model and some unanswered questions. And finally we throw caution to the wind and state what we think are some philosophical implications of this descriptive model, one of which is that neither normative ethics nor metaethics can be done behind a firewall. There can be little valid ethical inquiry that is not anchored in the facts of a particular species, so moral philosophers had best get a good grasp of the empirical facts of moral psychology.

**Question 1: Where Do Moral Beliefs and Motivations Come From?**

When a magician shows us an empty hat and then pulls a rabbit out of it, we all know there is a trick. Somehow or other, the rabbit had to be put into the hat. Infants and toddlers certainly seem like empty hats as far as morality is concerned, and then, somehow, by the time they are teenagers, they have morality. How is this trick accomplished? There are three main families of answers: empiricist, rationalist, and moral sense theories.

Most theories, lay and academic, have taken an empiricist approach. As with the magician’s rabbit, it just seems obvious that morality must have come from outside in. People in many cultures have assumed that God is the magician, revealing moral laws to people by way of prophets and divinely appointed kings. People are supposed to learn the laws and then follow them. The idea that morality is internalized is made most concrete in the Old Testament, in which Adam and Eve literally ingest morality when they bite into the forbidden fruit. When God finds out they have eaten of the “tree of the knowledge of good and evil,” he says “behold, the man has become like one of us, knowing good and evil” (Genesis, 3:22).

In the twentieth century, most people who didn’t buy the God theory bought a related empiricist, blank-slate, or “empty-hat” model: morality comes from society (which Durkheim said was God anyway), via the media and parents. For the behaviorists, morality was any set of responses that society happened to reward (Skinner, 1971). For Freud (1976/1900), morality comes from the father when a boy resolves his oedipal complex by internalizing the father’s superego. Some modern parents fear that morality comes from the barrage of images and stories their children see on TV. However, true blank-slate theories began to die when Garcia and Koelling (1966) demonstrated that equipotentiality—the equal ability of any
response to get hooked up to any stimulus—was simply not true. It is now universally accepted in psychology that some things are easy to learn (e.g., fearing snakes), while others (fearing flowers or hating fairness) are difficult or impossible. Nobody in psychology today admits to believing in the blank slate, although as Pinker (2002) has shown, in practice many psychologists stay as close to the blank slate as they can, often closer than the evidence allows.

The main alternative to empiricism has long been rationalism—the idea that reason plays a dominant role in our attempt to gain knowledge. Rationalists such as Descartes usually allow for the existence of innate ideas (such as the idea of God or perfection) and for the importance of sense perceptions, but they concentrate their attention on the processes of reasoning and inference by which people can extend their knowledge with certainty outwards from perceptions and innate ideas. Rationalist approaches to morality usually posit relatively little specific content—perhaps a few a priori concepts such as noncontradiction, or harm, or ought. The emphasis instead is on the act of construction, on the way that a child builds up her own moral understanding, and her ability to justify her judgments, as her developing mind with its all-purpose information processor becomes more and more powerful. Piaget, for example, allowed that children feel sympathy when they see others suffer. He then worked out the way the child gradually comes to understand and respect rules that help children get along, share, and thereby reduce suffering. “All morality consists in a system of rules, and the essence of all morality is to be sought for in the respect which the individual acquires for these rules” (Piaget, 1965/1932, p. 13).

Lawrence Kohlberg (1969, 1971) built on the foundation Piaget had laid to create the best known theory of moral development. In Kohlberg’s theory, young children are egocentric and concrete; they think that right and wrong are determined by what gets rewarded and punished. However, as their cognitive abilities mature around the ages of 6 to 8 and they become able to “decenter,” to look at situations through the eyes of others, they come to appreciate the value of rules and laws. As their abstract reasoning abilities mature around puberty, they become able to think about the reasons for having laws, and about how to respond to laws that are unjust. Cognitive development, however, is just a prerequisite for moral development; it does not create moral progress automatically. For moral progress to occur, children need plenty of “role-taking opportunities,” such as working out disputes during playground games or taking part in student government. Kohlberg’s approach to moral development was inspiring to
many people in the 1960s and 1970s, for it presented a picture of an active child, creating morality for herself, not just serving as a passive receptacle for social conditioning. Elliot Turiel (1983) continued this work, showing how children figure out that different kinds of rules and practices have different statuses. Moral rules, which are about harm, rights, and justice, have a different foundation and are much less revisable than social-conventional rules, which in turn are different from personal rules. As adults throw rule after rule at children, the children sort the rules into different cognitive bins (domains of social knowledge) and then figure out for themselves how and when to use—or reject—the different kinds of rules.

To give you a sense of a rationalist approach, we report the transcript of a remarkable interview that one of us (J. H.) overheard about the origin of moral rules. The interview was conducted in the bathroom of a McDonald’s restaurant in northern Indiana. The person interviewed—the subject—was a Caucasian male roughly 30 years old. The interviewer was a Caucasian male approximately 4 years old. The interview began at adjacent urinals:

Interviewer: Dad, what would happen if I pooped in here [the urinal]?
Subject: It would be yucky. Go ahead and flush. Come on, let's go wash our hands.

[The pair then moved over to the sinks]
Interviewer: Dad, what would happen if I pooped in the sink?
Subject: The people who work here would get mad at you.

Interviewer: What would happen if I pooped in the sink at home?
Subject: I'd get mad at you.

Interviewer: What would happen if YOU pooped in the sink at home?
Subject: Mom would get mad at me.

Interviewer: Well, what would happen if we ALL pooped in the sink at home?
Subject: [pause . . .] I guess we'd all get in trouble.

Interviewer: [laughing] Yeah, we'd all get in trouble!

Subject: Come on, let's dry our hands. We have to go.

If we analyze this transcript from a Kohlbergian perspective, the subject appears to score at the lowest stage: things seem to be wrong because they are punished. However, note the skill and persistence of the interviewer, who probes for a deeper answer by changing the transgression to remove a punishing agent. Yet even when everyone cooperates in the rule violation so that nobody can play the role of punisher, the subject still clings to a notion of cosmic or immanent justice in which, somehow, the whole family would “get in trouble.”
Of course, we didn’t really present this transcript to illustrate the depth and subtlety of Kohlberg’s approach. (For such an overview, see Lapsley, 1996; Kurtines & Gewirtz, 1995.) We presented it to show a possible limitation, in that Kohlberg and Turiel paid relatively little attention to the emotions. In each of his statements, the father is trying to socialize his curious son by pointing to moral emotions. He tries to get his son to feel that pooping in urinals and sinks is wrong. Disgust and anger (and the other moral emotions) are watchdogs of the moral world (Haidt, 2003; Rozin, Lowery, Imada, & Haidt, 1999), and we believe they play a very important role in moral development. This brings us to the third family of approaches: moral sense theories.

When God began to recede from scientific explanations in the sixteenth century, some philosophers began to wonder if God was really needed to explain morality either. In the seventeenth and eighteenth centuries, English and Scottish philosophers such as the third Earl of Shaftesbury, Frances Hutcheson, and Adam Smith surveyed human nature and declared that people are innately sociable, and that they are both benevolent and selfish. However it was David Hume (1775/1777) who worked out the details and implications of this approach most fully:

There has been a controversy started of late . . . concerning the general foundation of Morals; whether they be derived from Reason, or from Sentiment; whether we attain the knowledge of them by a chain of argument and induction, or by an immediate feeling and finer internal sense; whether, like all sound judgments of truth and falsehood, they should be the same to every rational intelligent being; or whether, like the perception of beauty and deformity, they be founded entirely on the particular fabric and constitution of the human species. (p. 2)

We added the italics above to show which side Hume was on. This passage is extraordinary for two reasons. First, it is a succinct answer to Question 1: Where do moral beliefs and motivations come from? They come from sentiments which give us an immediate feeling of right or wrong, and which are built into the fabric of human nature. Hume’s answer to Question 1 is our answer too, and much of the rest of our essay is an elaboration of this statement, using evidence and theories that Hume did not have available to him. However, this statement is also extraordinary as a statement about the controversy “started of late.” Hume’s statement is just as true in 2007 as it was in 1776. There really is a controversy started of late (in the 1980s), a controversy between rationalist approaches (based on Piaget and Kohlberg) and moral sense or intuitionist theories (e.g., Kagan, 1984; Frank, 1988; Haidt, 2001; Shweder & Haidt, 1993; J. Q. Wilson, 1993). We will not try to be fair and unbiased guides to this debate (indeed, our theory
says you should not believe us if we tried to be). Instead, we will make the case for a moral sense approach to morality, based on a small set of innately prepared, affectively valenced moral intuitions. We will contrast this approach to a rationalist approach, and we will refer the reader to other views when we discuss limitations of our approach. The contrast is not as stark as it seems: the SIM includes reasoning at several points, and rationalist approaches often assume some innate moral knowledge, but there is a big difference in emphasis. Rationalists say the real action is in reasoning; intuitionists say it’s in quick intuitions, gut feelings, and moral emotions.

Question 2: How Does Moral Judgment Work?

Brains evaluate and react. They are clumps of neural tissue that integrate information from the external and internal environments to answer one fundamental question: approach or avoid? Even one-celled organisms must answer this question, but one of the big selective advantages of growing a brain was that it could answer the question better and then initiate a more finely tailored response.

The fundamental importance of the good–bad or approach–avoid dimension is one of the few strings that runs the entire length of modern psychology. It was present at the birth, when Wilhelm Wundt (1907, as quoted by Zajonc, 1980) formulated the doctrine of “affective primacy,” which stated that the affective elements of experience (like–dislike, good–bad) reach consciousness so quickly and automatically that we can be aware of liking something before we know what it is. The behaviorists made approach and avoidance the operational definitions of reward and punishment, respectively. Osgood (1962) found that evaluation (good–bad) was the most basic dimension of all judgments. Zajonc (1980) argued that the human mind is composed of an ancient, automatic, and very fast affective system and a phylogenetically newer, slower, and motivationally weaker cognitive system. Modern social cognition research is largely about the disconnect between automatic processes, which are fast and effortless, and controlled processes, which are slow, conscious, and heavily dependent on verbal thinking (Bargh & Ferguson, 2000; Chaiken & Trope, 1999; Wegner & Bargh, 1998).

The conclusion at the end of this string is that the human mind is always evaluating, always judging everything it sees and hears along a “good–bad” dimension (see Kahneman, 1999). It doesn’t matter whether we are looking at men’s faces, lists of appetizers, or Turkish words; the brain has a kind of
gauge (sometimes called a “like-ometer”) that is constantly moving back and forth, and these movements, these quick judgments, influence whatever comes next. The most dramatic demonstration of the like-ometer in action is the recent finding that people are slightly more likely than chance to marry others whose first name shares its initial letter with their own, they are more likely to move to cities and states that resemble their names (Phil moves to Philadelphia; Louise to Louisiana), and they are more likely to choose careers that resemble their names (Dennis finds dentistry more appealing; Lawrence is drawn to law; Pelham, Mirenberg, & Jones, 2002). Quick flashes of pleasure, caused by similarity to the self, make some options “just feel right.”

This perspective on the inescapably affective mind is the foundation of the SIM, presented in figure 4.1 (from Haidt, 2001). The model is composed of six links, or psychological processes, which describe the relationships among an initial intuition of good versus bad, a conscious moral judgment, and conscious moral reasoning. The first four links are the core of the model, intended to capture the great majority of judgments for most

![Figure 4.1](image)

**Figure 4.1**

The social intuitionist model of moral judgment. The numbered links, drawn for Person A only, are (1) the intuitive judgment link, (2) the post hoc reasoning link, (3) the reasoned persuasion link, and (4) the social persuasion link. Two additional links are hypothesized to occur less frequently, (5) the reasoned judgment link and (6) the private reflection link. (Reprinted from Haidt, 2001)
people. Links 5 and 6 are hypothesized to occur rarely but should be of
great interest to philosophers because they are used to solve dilemmas and
because philosophers probably use these links far more than most people
(Kuhn, 1991). The existence of each link as a psychological process is well
supported by research, presented below. However, whether everyday moral
judgment is best captured by this particular arrangement of processes is
still controversial (Greene, volume 3 of this collection; Pizarro & Bloom,
2003), so the SIM should be considered a hypothesis for now, rather than
an established fact. The model and a brief description of the six links are
presented next.

**Link 1: The Intuitive Judgment Link**
The SIM is founded on the idea that moral judgment is a ubiquitous
product of the ever-evaluating mind. Like aesthetic judgments, moral judg-
ments are made quickly, effortlessly, and intuitively. We see an act of vio-
ence, or hear about an act of gratitude, and we experience an instant flash
of evaluation, which may be as hard to explain as the affective response
to a face or a painting. That's the intuition. “Moral intuition” is defined
as the sudden appearance in consciousness, or at the fringe of conscious-
ness, of an evaluative feeling (like–dislike, good–bad) about the character
or actions of a person, without any conscious awareness of having gone
through steps of search, weighing evidence, or inferring a conclusion
(modified from Haidt, 2001, p. 818). This is the “finer internal sense” that
Hume talked about. In most cases this flash of feeling will lead directly to
the conscious condemnation (or praise) of the person in question, often
including verbal thoughts such as “What a bastard” or “Wow, I can’t
believe she’s doing this for me!” This conscious experience of blame or
praise, including a belief in the rightness or wrongness of the act, is the
moral judgment. Link 1 is the tight connection between flashes of intu-
ition and conscious moral judgments. However this progression is not
inevitable: often a person has a flash of negative feeling, for example,
toward stigmatized groups (easily demonstrated through implicit mea-
surement techniques such as the Implicit Association Test; Greenwald,
McGhee, & Schwartz, 1998), yet because of one’s other values, one resists
or blocks the normal tendency to progress from intuition to consciously
endorsed judgment.

These flashes of intuition are not dumb; as with the superb mental soft-
ware that runs visual perception, they often hide a great deal of sophisti-
cated processing occurring behind the scenes. Daniel Kahneman, one of
the leading researchers of decision making, puts it this way:
We become aware only of a single solution—this is a fundamental rule in perceptive processing. All other solutions that might have been considered by the system—and sometimes we know that alternative solutions have been considered and rejected—we do not become aware of. So consciousness is at the level of a choice that has already been made. (quoted in Jaffe, 2004, p. 26)

Even if moral judgments are made intuitively, however, we often feel a need to justify them with reasons, much more so than we do for our aesthetic judgments. What is the relationship between the reasons we give and the judgments we reach?

**Link 2: The Post Hoc Reasoning Link**

Studies of reasoning describe multiple steps, such as searching for relevant evidence, weighing evidence, coordinating evidence with theories, and reaching a decision (Kuhn, 1989; Nisbett & Ross, 1980). Some of these steps may be performed unconsciously, and any of the steps may be subject to biases and errors, but a key part of the definition of reasoning is that it has steps, at least two of which are performed consciously. Galotti (1989, p. 333), in her definition of everyday reasoning, specifically excludes “any one-step mental processes” such as sudden flashes of insight, gut reactions, and other forms of “momentary intuitive response.” Building on Galotti (1989), moral reasoning can be defined as conscious mental activity that consists of transforming given information about people in order to reach a moral judgment (Haidt, 2001, p. 818). To say that moral reasoning is a conscious process means that the process is intentional, effortful, and controllable and that the reasoner is aware that it is going on (Bargh, 1994).

The SIM says that moral reasoning is an effortful process (as opposed to an automatic process), usually engaged in after a moral judgment is made, in which a person searches for arguments that will support an already-made judgment. This claim is consistent with Hume’s famous claim that reason is “the slave of the passions, and can pretend to no other office than to serve and obey them” (Hume, 1969/1739, p. 462). Nisbett and Wilson (1977) demonstrated such post hoc reasoning for causal explanations. When people are tricked into doing a variety of things, they readily make up stories to explain their actions, stories that can often be shown to be false. People often know more than they can tell, but when asked to introspect on their own mental processes, people are quite happy to tell more than they can know, expertly crafting plausible-sounding explanations from a pool of cultural theories about why people generally do things (see Wilson, 2002, on the limits of introspection).
The most dramatic cases of post hoc confabulation come from Gazzaniga’s studies of split-brain patients (described in Gazzaniga, 1985). When a patient performs an action caused by a stimulus presented to the right cerebral hemisphere (e.g., getting up and walking away), the left hemisphere, which controls language, does not say “Hey, I wonder why I’m doing this!” Rather, it makes up a reason, such as “I’m going to get a soda.” Gazzaniga refers to the brain areas that provide a running post hoc commentary on our behavior as the “interpreter module.” He says that our conscious verbal reasoning is in no way the command center of our actions; it is rather more like a press secretary, whose job is to offer convincing explanations for whatever the person happens to do. Subsequent research by Kuhn (1991), Kunda (1990), and Perkins, Farady, and Bushey (1991) has found that everyday reasoning is heavily marred by the biased search only for reasons that support one’s already-favored hypothesis. People are extremely good at finding reasons for whatever they have done, are doing, or want to do in the future. In fact, this human tendency to search only for reasons and evidence on one side of a question is so strong and consistent in the research literature that it might be considered the chief obstacle to good thinking.

**Link 3: The Reasoned Persuasion Link**

The glaring one-sidedness of everyday human reasoning is hard to understand if you think that the goal of reasoning is to reach correct conclusions or to create accurate representations of the social world. However, many thinkers, particularly in evolutionary psychology, have argued that the driving force in the evolution of language was not the value of having an internal truth-discovering tool; it was the value of having a tool to help a person track the reputations of others, and to manipulate those others by enhancing one’s own reputation (Dunbar, 1996). People are able to reuse this tool for new purposes, including scientific or philosophical inquiry, but the fundamentally social origins of speech and internal verbal thought affect our other uses of language.

Links 3 and 4 are the social part of the SIM. People love to talk about moral questions and violations, and one of the main topics of gossip is the moral and personal failings of other people (Dunbar, 1996; Hom & Haidt, in preparation). In gossip people work out shared understandings of right and wrong, they strengthen relationships, and they engage in subtle or not-so-subtle acts of social influence to bolster the reputations of themselves and their friends (Hom & Haidt, in preparation; Wright, 1994). Allan Gibbard (1990) is perhaps the philosopher who is most sensitive to the
social nature of moral discourse. Gibbard took an evolutionary approach to this universal human activity and asked about the functions of moral talk. He concluded that people are designed to respond to what he called “normative governance,” or a general tendency to orient their actions with respect to shared norms of behavior worked out within a community. However, Gibbard did not assume that people blindly follow whatever norms they find; rather, he worked out the ways in which people show a combination of firmness in sticking to the norms that they favor, plus persuadability in being responsive to good arguments produced by other people. People strive to reach consensus on normative issues within their “parish,” that is, within the community they participate in. People who can do so can reap the benefits of coordination and cooperation. Moral discourse therefore serves an adaptive biological function, increasing the fitness of those who do it well.

Some evolutionary thinkers have taken this adaptive view to darker extremes. In an eerie survey of moral psychology, Robert Wright (1994) wrote:

The proposition here is that the human brain is a machine for winning arguments, a machine for convincing others that its owner is in the right—and thus a machine for convincing its owner of the same thing. The brain is like a good lawyer: given any set of interests to defend, it sets about convincing the world of their moral and logical worth, regardless of whether they in fact have any of either. Like a lawyer, the human brain wants victory, not truth. (p. 280)

This may offend you. You may feel the need to defend your brain’s honor. But the claim here is not that human beings can never think rationally or that we are never open to new ideas. Lawyers can be very reasonable when they are off duty, and human minds can be too. The problem comes when we find ourselves firmly on one side of a question, either because we had an intuitive or emotional reaction to it or because we have interests at stake. It is in those situations, which include most acts of moral judgment, that conscious verbal moral reasoning does what it may have been designed to do: argue for one side.

It is important to note that “reasoned persuasion” does not necessarily mean persuasion via logical reasons. The reasons that people give to each other are best seen as attempts to trigger the right intuitions in others. For example, here is a quotation from an activist arguing against the practice, common in many cultures, of altering the genitalia of both boys and girls either at birth or during initiation rites at puberty: “This is a clear case of child abuse. It’s a form of reverse racism not to protect these girls from barbarous practices that rob them for a lifetime of their God-given right to an
intact body” (Burstyn, 1995). These two sentences contain seven arguments against altering female genitalia, each indicated in italics. But note that each argument is really an attempt to frame the issue so as to push an emotional button, triggering seven different flashes of intuition in the listener. Rhetoric is the art of pushing the ever-evaluating mind over to the side the speaker wants it to be on, and affective flashes do most of the pushing.

Link 4: The Social Persuasion Link
There are, however, means of persuasion that don’t involve giving reasons of any kind. The most dramatic studies in social psychology are the classic studies showing just how easily the power of the situation can make people do and say extraordinary things. Some of these studies show obedience without persuasion (e.g., Milgram’s, 1963, “shock” experiments); some show conformity without persuasion (e.g., Asch’s, 1956, line-length experiments). But many show persuasion. Particularly when there is ambiguity about what is happening, people look to others to help them interpret what is going on and what they should think about what is going on. Sherif (1935) asked people to guess at how far a point of light was moving, back and forth. On this purely perceptual task, people were strongly influenced by their partner’s ratings. Latane and Darley (1970) put people into ambiguous situations where action was probably—but not definitely—called for, and the presence of another person who was unresponsive influenced people’s interpretations of and responses to potential emergencies. In study after classic study, people adjust their beliefs to fit with the beliefs of others, not just because they assume others have useful information but largely for the simple reason that they interact with these others, or even merely expect to interact (Darley & Berscheid, 1967). Recent findings on the “chameleon effect” show that people will automatically and unconsciously mimic the postures, mannerisms, and facial expressions of their interaction partners and that such mimicry leads the other person to like the mimicker more (Chartrand & Bargh, 1999).

Human beings are almost unique among mammals in being “ultrasocial”—that is, living in very large and highly cooperative groups of thousands of individuals, as bees and ants do (Richerson & Boyd, 1998). The only other ultrasocial mammals are the naked mole rats of East Africa, but they, like the bees and the ants, accomplish their ultrasociality by all being siblings and reaping the benefits of kin altruism. Only human beings cooperate widely and intensely with nonkin, and we do it in part through a set of social psychological adaptations that make us extremely sensitive to and influenceable by what other people think and feel. We have an intense
need to belong and to fit in (Baumeister & Leary, 1995), and our moral judgments are strongly shaped by what others in our “parish” believe, even when they don’t give us any reasons for their beliefs. Link 4, the social persuasion link, captures this automatic unconscious influence process.

These four links form the core of the SIM. The core of the model gives moral reasoning a causal role in moral judgment, but only when reasoning runs through other people. If moral reasoning is transforming information to reach a moral judgment, and if this process proceeds in steps such as searching for evidence and then weighing the evidence, then a pair of people discussing a moral issue meets the definition of reasoning. Reasoning, even good reasoning, can emerge from a dyad even when each member of the dyad is thinking intuitively and reasoning post hoc. As long as people are at least a little bit responsive to the reasons provided by their partners, there is the possibility that the pair will reach new and better conclusions than either could have on her own. People are very bad at questioning their own initial assumptions and judgments, but in moral discourse other people do this for us. To repeat: moral judgment should be studied as a social process, and in a social context moral reasoning matters.

Can a person ever engage in open-minded, non–post hoc moral reasoning in private? Yes. The loop described by the first four links in the SIM is intended to capture the great majority of moral judgments made by the great majority of people. However, many people can point to times in their lives when they changed their minds on a moral issue just from mulling the matter over by themselves or to dilemmas that were so well balanced that they had to reason things out. Two additional links are included to account for these cases, hypothesized to occur somewhat rarely outside of highly specialized subcultures such as that of philosophy, which provides years of training in unnatural modes of human thought.

**Link 5: The Reasoned Judgment Link**

People may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition. In such cases reasoning truly is causal and cannot be said to be the “slave of the passions.” However such reasoning is hypothesized to be rare, occurring primarily in cases in which the initial intuition is weak and processing capacity is high. In cases where the reasoned judgment conflicts with a strong intuitive judgment, a person will have a “dual attitude” (Wilson, Lindsey, & Schooler, 2000) in which the reasoned judgment may be expressed verbally, yet the intuitive judgment
continues to exist under the surface, discoverable by implicit measures such as the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998).

Philosophers have long tried to derive coherent and consistent moral systems by reasoning out from first principles. However when these reasoned moral systems violate people’s other moral intuitions, the systems are usually rejected or resisted. For example, Peter Singer’s (1979) approach to bioethical questions is consistent and humane in striving to minimize the suffering of sentient beings, but it leads to the logical conclusion that the life of a healthy chimpanzee deserves greater protection than that of an acephalic human infant who will never have consciousness. Singer’s work is a paragon of reasoned judgment, but because his conclusions conflict with many people’s inaccessible and unrevisable moral intuitions about the sanctity of human life, Singer is sometimes attacked by political activists and compared, absurdly, to the Nazis. (See also Derek Parfit’s, 1984, “repugnant conclusion” that we should populate the world much more fully, and Kant’s, 1969/1785, conclusion that one should not tell a lie to save the life of an innocent person.)

Link 6: The Private Reflection Link
In the course of thinking about a situation, a person may spontaneously activate a new intuition that contradicts the initial intuitive judgment. The most widely discussed method of triggering new intuitions is role taking (Selman, 1971). Simply by putting yourself into the shoes of another person you may instantly feel pain, sympathy, or other vicarious emotional responses. This is one of the principle pathways of moral reflection according to Piaget, Kohlberg, and other cognitive developmentalists. A person comes to see an issue or dilemma from more than one side and thereby experiences multiple competing intuitions. The final judgment may be determined either by going with the strongest intuition, or by using reasoning to weigh pros and cons or to apply a rule or principle (e.g., one might think “Honesty is the best policy”). This pathway amounts to having an inner dialogue with oneself (Tappan, 1997), obviating the need for a discourse partner. Is this really reasoning? As long as part of the process occurs in steps, in consciousness, it meets the definition given above for moral reasoning. However, all cases of moral reasoning probably involve a great deal of intuitive processing. William James described the interplay of reason and intuition in private deliberations as follows:

Reason, per se, can inhibit no impulses; the only thing that can neutralize an impulse is an impulse the other way. Reason may, however, make an inference which will excite the imagination so as to set loose the impulse the other way; and thus,
though the animal richest in reason might also be the animal richest in instinctive impulses, too, he would never seem the fatal automaton which a merely instinctive animal would be. (quoted in Ridley, 2004, p. 39)

James suggests that what feels to us like reasoning is really a way of helping intuition (impulse, instinct) to do its job well: we consider various issues and entailments of a decision and, in the process, allow ourselves to feel our way to the best answer using a combination of conscious and unconscious, affective and “rational” processes. This view fits the findings of Damasio (1994) that reasoning, when stripped of affective input, becomes inept. Reasoning requires affective channeling mechanisms. The private reflection link describes this process, in which conflicts get worked out in a person’s mind without the benefit of social interaction. It is a kind of reasoning (it involves at least two steps), yet it is not the kind of reasoning described by Kohlberg and the rationalists.

Private reflection is necessary whenever intuitions conflict, or in those rare cases where a person has no intuition at all (such as on some public policy issues where one simply does not know enough to have an opinion). Conflicting intuitions may be fairly common, particularly in moral judgment problems that are designed specifically to be dilemmas. Greene (volume 3 of this collection), for example, discusses the “crying baby” problem, in which if you do not smother your child, the child’s cries will alert the enemy soldiers searching the house, which will lead in turn to the deaths of you, your child, and the other townspeople hiding in the basement. Gut feelings say “no, don’t kill the child,” yet as soon as one leans toward making the “no” response, one must deal with the consequence that the choice leads to death for many people, including the baby. Greene’s fMRI data show that, in these difficult cases in particular, the dorsolateral prefrontal cortex is active, indicating “cooler” reasoning processes at work. But does a slow “yes” response indicate the victory of the sort of reasoning a philosopher would respect over dumb emotional processes? We think such cases are rather the paradigm of the sort of affective reasoning that James and Damasio described: there is indeed a conflict between potential responses, and additional areas of the brain become active to help resolve this conflict, but ultimately the person decides based on a feeling of rightness, rather than a deduction of some kind.

If you would like to feel these affective channeling mechanisms in action, just look at slavery in the American South from a slaveholder’s point of view, look at Auschwitz from Hitler’s point of view, or look at the 9/11 attacks from Bin Laden’s point of view. There are at least a few supportive reasons on the “other” side in each case, but it will probably cause
you pain to examine those reasons and weigh the pros and cons. It is as though our moral deliberations are structured by the sorts of invisible fences that keep suburban dogs from straying over property lines, giving them an electric shock each time they get too near a border. If you are able to rise to this challenge, if you are able to honestly examine the moral arguments in favor of slavery and genocide (along with the much stronger arguments against them), then you are likely to be either a psychopath or a philosopher. Philosophers are one of the only groups that have been found spontaneously to look for reasons on both sides of a question (Kuhn, 1991); they excel at examining ideas “dispassionately.”

**Question 3: Why Should You Believe Us?**

Our most general claim is that the action in morality is in the intuitions, not in reasoning. Our more specific claim is that the SIM captures the interaction between intuition, judgment, and reasoning. What is the evidence for these claims? In this section we briefly summarize the findings from relevant empirical studies.

**Moral Judgment Interviews**

In the 1980s a debate arose between Elliot Turiel (1983), who said that the moral domain is universally limited to issues of harm, rights, and justice, and Richard Shweder (Shweder, Mahapatra, & Miller, 1987), who said that the moral domain is variable across cultures. Shweder et al. showed that in Orissa, India, the moral domain includes many issues related to food, clothing, sex roles, and other practices Turiel would label as “social conventions.” However Turiel, Killen, and Helwig (1987) argued that most of Shweder’s research vignettes contained harm, once you understand how Indians construed the violations.

Haidt, Koller, and Dias (1993) set out to resolve this debate by using a class of stories that had not previously been used: harmless taboo violations. They created a set of stories that would cause an immediate affective reaction in people but that upon reflection would be seen to be harmless and unrelated to issues of rights or justice. For example, a family eats its pet dog after the dog was killed by a car; a woman cuts up an old flag to create rags with which to clean her toilet; a man uses a chicken carcass for masturbation, and afterwards he cooks and eats the carcass. These stories were presented to 12 groups of subjects (360 people in all) during interviews modeled after Turiel (1983). Half of the subjects were adults, and half were children (ages 10–12; they did not receive the chicken.
story); half were of high social class, and half of low; and they were residents of three cities: Recife, Brazil; Porto Alegre, Brazil; and Philadelphia, U.S.A. The basic finding was that the high social class adult groups, which were composed of college students, conformed well to Turiel’s predictions. They treated harmless taboo violations as strange and perhaps disgusting, but not morally wrong. They said, for example, that such behaviors would not be wrong in another culture where they were widely practiced. The other groups, however, showed the broader moral domain that Shweder had described. They overwhelmingly said that these actions were wrong and universally wrong, even as they explicitly stated that nobody was harmed. They treated these acts as moral violations, and they justified their condemnation not by pointing to victims but by pointing to disgust or disrespect, or else by pointing simply to norms and rules (“You just don’t have sex with a chicken!”). College students largely limited themselves to a mode of ethical discourse that Shweder, Much, Mahapatra, and Park (1997) later called the “ethics of autonomy” (judgments relating to issues of harm, rights, and justice), while the other groups showed a much broader moral domain including the “ethics of community” (issues of respect, duty, hierarchy, and group obligation) and to a lesser extent the “ethics of divinity” (issues of purity, sanctity, and recognition of divinity in each person).

While conducting these interviews, however, Haidt noticed an interesting phenomenon: most subjects gave their initial evaluations almost instantly, but then some struggled to find a supporting reason. For example, a subject might say, hesitantly, “It’s wrong to eat your dog because . . . you might get sick.” When the interviewer pointed out that the dog meat was fully cooked and so posed no more risk of illness than any other meat, subjects rarely changed their minds. Rather, they searched harder for additional reasons, sometimes laughing and confessing that they could not explain themselves. Haidt and Hersh (2001) noticed the same thing in a replication study that asked political liberals and conservatives to judge a series of harmless sexual behaviors, including various forms of masturbation, homosexuality, and consensual incest. Haidt and Hersh called this state of puzzled inability to justify a moral conviction “moral dumbfounding.”

We (Haidt, Bjorklund, & Murphy, 2000) brought moral dumbfounding into the lab to examine it more closely. In Study 1 we gave subjects five tasks: Kohlberg’s Heinz dilemma (should Heinz steal a drug to save his wife’s life?), which is known to elicit moral reasoning; two harmless taboo violations (consensual adult sibling incest and harmless cannibalism of an
unclaimed corpse in a pathology lab); and two behavioral tasks that were designed to elicit strong gut feelings: a request to sip a glass of apple juice into which a sterilized dead cockroach had just been dipped and a request to sign a piece of paper that purported to sell the subject’s soul to the experimenter for $2 (the form explicitly said that it was not a binding contract, and the subject was told she could rip up the form immediately after signing it). The experimenter presented each task and then played devil’s advocate, arguing against anything the subject said. The key question was whether subjects would behave like (idealized) scientists, looking for the truth and using reasoning to reach their judgments, or whether they would behave like lawyers, committed from the start to one side and then searching only for evidence to support that side, as the SIM suggests.

Results showed that on the Heinz dilemma people did seem to use some reasoning, and they were somewhat responsive to the counterarguments given by the experimenter. (Remember the social side of the SIM: people are responsive to reasoning from another person when they do not have a strong countervailing intuition.) However, responses to the two harmless taboo violations were more similar to responses on the two behavioral tasks: very quick judgment was followed by a search for supporting reasons only; when these reasons were stripped away by the experimenter, few subjects changed their minds, even though many confessed that they could not explain the reasons for their decisions. In Study 2 we repeated the basic design while exposing half of the subjects to a cognitive load—an attention task that took up some of their conscious mental work space—and found that this load increased the level of moral dumbfounding without changing subjects’ judgments or their level of persuadability.

Manipulating Intuitions

In other studies we have directly manipulated the strength of moral intuitions without changing the facts being judged, to test the prediction that Link 1 (intuitive judgment) directly causes, or at least influences, moral judgments. Wheatley and Haidt (2005) hypnotized one group of subjects to feel a flash of disgust whenever they read the word “take”; another group was hypnotized to feel disgust at the word “often.” Subjects then read six moral judgment stories, each of which included either the word “take” or the word “often.” Only highly hypnotizable subjects who were amnesic for the posthypnotic suggestion were used. In two studies, the flash of disgust that subjects felt while reading three of their six stories made their moral judgments more severe. In Study 2, a seventh story was included in which there was no violation whatsoever, to test the limits of the phe-
nomenon: “Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take] <often picks> topics that appeal to both professors and students in order to stimulate discussion.” We predicted that with no violation of any kind, subjects would be forced to override their feelings of disgust, and most did. However, one third of all subjects who encountered their disgust word in the story still rated Dan’s actions as somewhat morally wrong, and several made up post hoc confabulations reminiscent of Gazzaniga’s findings. One subject justified his condemnation of Dan by writing “it just seems like he’s up to something.” Another wrote that Dan seemed like a “popularity seeking snob.” These cases provide vivid examples of reason playing its role as slave to the passions.

In another experiment, Bjorklund and Haidt (in preparation) asked subjects to make moral judgments of norm violation scenarios that involved disgusting features. In order to manipulate the strength of the intuitive judgment made in Link 1, one group of subjects got a version of the scenarios where the disgusting features were vividly described, and another group got a version where they were not vividly described. Subjects who got scenarios with vividly described disgust made stronger moral judgments, even though the disgusting features were morally irrelevant.

Another way of inducing irrelevant disgust is to alter the environment in which people make moral judgments. Schnall, Haidt, Clore, and Jordan (2007) asked subjects to make moral judgments while seated either at a clean and neat desk or at a dirty desk with fast food wrappers and dirty tissues strewn about. The dirty desk was assumed to induce low-level feelings of disgust and avoidance motivations. Results showed that the dirty desk did make moral judgments more severe, but only for those subjects who had scored in the upper half of a scale measuring “private body consciousness,” which means the general tendency to be aware of bodily states and feelings such as hunger and discomfort. For people who habitually listen to their bodies, extraneous feelings of disgust did affect moral judgment.

**Neuroscientific Evidence**

A great deal of neuroscience research supports the idea that flashes of affect are essential for moral judgment (see Greene & Haidt, 2002, for a review). Damasio’s (1994) work on “acquired sociopathy” shows that damage to the ventromedial prefrontal cortex, an area that integrates affective responses with higher cognition, renders a person morally incompetent, particularly if the damage occurs in childhood (Anderson et al., 1999), suggesting that
emotions are necessary for moral learning. When emotion is removed from decision making, people do not become hyperlogical and hyperethical; they become unable to feel the rightness and wrongness of simple decisions and judgments. Joshua Greene and his colleagues at Princeton have studied the brains of healthy people making moral judgments while in an fMRI scanner (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). They found that the distinctions people make between various classes of moral dilemmas are predicted by whether or not certain brain areas involved in emotional responding are more active. When considering dilemmas with direct physical harm (e.g., pushing one man off of a footbridge to stop a trolley from killing five men), most people have a quick flash of activity in the medial prefrontal cortex and then say that it is not permissible to do this. They make deontological judgments, which they justify with references to rights or to moral absolutes. When they think about cases where the harm is less direct but the outcome is the same (e.g., throwing a switch to shift the train from killing five to killing one), they have no such flash, and most people choose the utilitarian response.

Greene (volume 3 of this collection) makes the provocative argument that deontological judgments are really just gut feelings dressed up with fancy sounding justifications. He suggests that such judgments conform closely to the mechanisms described by the SIM. However, in cases where people override their gut feelings and choose the utilitarian response, Greene believes that the SIM may fail to capture the causal role of moral reasoning. Greene bases this suggestion on evidence of “cognitive conflict” in moral judgment (Greene, Nystrom, Engell, Darley, & Cohen, 2004). For example, people take longer to respond to difficult personal moral dilemmas, such as the crying baby story (described above), than to other types of stories, and when people do choose the utilitarian response to difficult dilemmas, they show a late surge of activity in the dorsolateral prefrontal cortex (suggesting “cognitive” activity), as well as increased activity in the anterior cingulate cortex (indicating a response conflict).

We see no contradiction between Greene’s results and the SIM. While some critics erroneously reduce the SIM to the claim that moral reasoning doesn’t happen or doesn’t matter (see Saltzstein & Kasachkoff, 2004), we point out that the SIM says that reasoning happens between people quite often (link 3), and within individuals occasionally (links 5 and 6). Furthermore, the SIM is not about “cognition” and “emotion”; it is about two kinds of cognition: fast intuition (which is sometimes but not always a part of an emotional response) and slow reasoning. Intuitions often conflict or lead to obviously undesirable outcomes (such as the death of everyone if
the crying baby is not smothered), and when they do, the conflict must get resolved somehow. This resolution requires time and the involvement of brain areas that handle response conflict (such as the anterior cingulate cortex). The private reflection link of the SIM is intended to handle exactly these sorts of cases in which a person considers responses beyond the initial intuitive response. According to the definition of reasoning given in the SIM, private reflection is a kind of reasoning—it involves at least two steps carried out in consciousness. However, this reasoning is not, we believe, the sort of logical and dispassionate reasoning that philosophers would respect; it is more like the kind of weighing of alternatives in which feelings play a crucial role (as described by James and Damasio, above). Hauser, Young, and Cushman’s finding (this volume) that people usually cannot give a good justification for their responses to certain kinds of trolley problems is consistent with our claim that the process of resolving moral dilemmas and the process of formulating justifications to give to other people are separate processes, even when both can be considered kinds of reasoning.

**Question 4: What Exactly Are the Intuitions?**

If we want to rebuild moral psychology on an intuitionist foundation, we had better have a lot more to say about what intuitions are, and about why people have the particular intuitions they have. We look to evolution to answer these questions. One could perfectly well be an empiricist intuitionist—one might believe that children simply develop intuitions or reactions for which they are reinforced; or one might believe that children have a general tendency to take on whatever values they see in their parents, their peers, or the media. Of course, social influence is important, and the social links of the SIM are intended to capture such processes. However, we see two strong arguments against a fully empiricist approach in which intuitions are entirely learned. The first, pointed out by Tooby, Cosmides, and Barrett (2005), is that children routinely resist parental efforts to get them to care about, value, or desire things. It is just not very easy to shape children, unless one is going with the flow of what they already like. It takes little or no work to get 8-year-old children to prefer candy to broccoli, to prefer being liked by their peers to being approved of by adults, or to prefer hitting back to loving their enemies. Socializing the reverse preferences would be difficult or impossible. The resistance of children to arbitrary or unusual socialization has been the downfall of many utopian efforts. Even if a charismatic leader can recruit a group of
unusual adults able to believe in universal love while opposing all forms of hatred and jealousy, nobody has ever been able to raise the next generation of children to take on such unnatural beliefs.

The second argument is that despite the obvious cultural variability of norms and practices, there is a small set of moral intuitions that is easily found in all societies, and even across species. An analogy to cuisine might be useful: human cuisines are cultural products, and each is unique—a set of main ingredients and plant-based flavorings that mark food as familiar and safe (Rozin, 1982). However, cuisines are built on top of an evolved sensory system including just five kinds of taste receptors on the tongue, plus a more complex olfactory system. The five kinds of taste buds have obvious adaptive benefits: sweetness indicates fruit and safety; bitterness indicates toxins and danger; glutamate indicates meat. The structure of the human tongue, nose, and brain place constraints on cuisines while leaving plenty of room for creativity. One could even say that the constraints make creativity possible, including the ability to evaluate one meal as better than another.

Might there be a small set of moral intuitions underlying the enormous diversity of moral “cuisines?” Just such an analogy was made by the Chinese philosopher Mencius 2,400 years ago:

There is a common taste for flavor in our mouths, a common sense for sound in our ears, and a common sense of beauty in our eyes. Can it be that in our minds alone we are not alike? What is it that we have in common in our minds? It is the sense of principle and righteousness. The sage is the first to possess what is common in our minds. Therefore moral principles please our minds as beef and mutton and pork please our mouths. (Mencius, quoted in Chan, 1963, p. 56).

Elsewhere Mencius specifies that the roots, or common principles of human morality are to be found in moral feelings such as commiseration, shame, respect, and reverence (Chan, 1963, p. 54).

Haidt and Joseph (2004) set out to list these common principles a bit more systematically, reviewing five works that were rich in detail about moral systems. Two of the works were written to capture what is universal about human cultures: Donald Brown’s (1991) catalogue *Human Universals* and Alan Fiske’s (1992) grand integrative theory of the four models of social relations. Two of the works were designed primarily to explain differences across cultures in morality: Schwartz and Bilsky’s (1990) widely used theory of 15 values, and Richard Shweder’s theory of the “big 3” moral ethics—autonomy, community, and divinity (Shweder et al., 1997). The fifth work was Frans de Waal’s (1996) survey of the roots or precursors of morality in other animals, primarily chimpanzees, *Good Natured*. We (Haidt
& Joseph) simply listed all the cases where some aspect of the social world was said to trigger approval or disapproval; that is, we tried to list all the things that human beings and chimpanzees seem to value or react to in the behavior of others. We then tried to group the elements that were similar into a smaller number of categories, and finally we counted up the number of works (out of five) that each element appeared in. The winners, showing up clearly in all five works, were harm/care (a sensitivity to or dislike of signs of pain and suffering in others, particularly in the young and vulnerable), fairness/reciprocity (a set of emotional responses related to playing tit-for-tat, such as negative responses to those who fail to repay favors), and authority/respect (a set of concerns about navigating status hierarchies, e.g., anger toward those who fail to display proper signs of deference and respect). We believe these three issues are excellent candidates for being the “taste buds” of the moral domain. In fact, Mencius specifically included emotions related to harm (commiseration) and authority (respect, and reverence) as human universals.

We tried to see how much moral work these three sets of intuitions could do and found that we could explain most but not nearly all of the moral virtues and concerns that are common in the world’s cultures. There were two additional sets of concerns that were widespread but that had only been mentioned in three or four of the five works: concerns about purity/sanctity (related to the emotion of disgust, necessary for explaining why so many moral rules relate to food, sex, menstruation, and the handling of corpses), and concerns about boundaries between in-group and out-group.² Liberal moral theorists may dismiss these concerns as matters of social convention (for purity practices) or as matters of prejudice and exclusion (for in-group concerns), but we believe that many or most cultures see matters of purity, chastity, in-group loyalty, and patriotism as legitimate parts of their moral domain (see Haidt & Graham, in press-a, in press-b).

We (Haidt, Joseph, & Bjorklund) believe these five sets of intuitions should be seen as the foundations of intuitive ethics. For each one, a clear evolutionary story can be told and has been told many times. We hope nobody will find it controversial to suppose that evolution has prepared the human mind to easily develop a sensitivity to issues related to harm/care, fairness/reciprocity, in-group/loyalty, and authority/respect. The only set of intuitions with no clear precursor in other animals is purity/sanctity. However, concerns about purity and pollution require the emotion of disgust and its cognitive component of contamination sensitivity, which only human beings older than the age of 7 have fully
mastered (Rozin, Fallon, & Augustoni-Ziskind, 1985). We think it is quite sensible to suppose that most of the foundations of human morality are many millions of years old, but that some aspects of human morality have no precursors in other animals.

Now that we have identified five promising areas or clusters of intuition, how exactly are they encoded in the human mind? There are a great many ways to think about innateness. At the mildest extreme is a general notion of “preparedness,” the claim that animals are prepared (by evolution) to learn some associations more easily than others (Seligman, 1971). For example, rats can more easily learn to associate nausea with a new taste than with a new visual stimulus (Garcia & Koelling, 1966), and monkeys (and humans) can very quickly acquire a fear of snakes from watching another monkey (or human) reacting with fear to a snake, but it is very hard to acquire a fear of flowers by such social learning (Mineka & Cook, 1988). The existence of preparedness as a product of evolution is uncontroversial in psychology. Everyone accepts at least that much writing on the slate at birth. Thus, the mildest version of our theory is that the human mind has been shaped by evolution so that children can very easily be taught or made to care about harm, fairness, in-groups, authority, and purity; however, children have no innate moral knowledge—just a preparedness to acquire certain kinds of moral knowledge and a resistance to acquiring other kinds (e.g., that all people should be loved and valued equally).

At the other extreme is the idea of the massively modular mind, championed by evolutionary psychologists such as Pinker (1997) and Cosmides and Tooby (1994). On this view, the mind is like a Swiss army knife with many tools, each one an adaptation to the long-enduring structure of the world. If every generation of human beings faced the threat of disease from bacteria and parasites that spread by physical touch, minds that had a contamination-sensitivity module built in (i.e., feel disgust toward feces and rotting meat and also toward anything that touches feces or rotting meat) were more likely to run bodies that went on to leave surviving offspring than minds that had to learn everything from scratch using only domain-general learning processes. As Pinker (2002) writes, with characteristic flair: “The sweetness of fruit, the scariness of heights, and the vileness of carrion are fancies of a nervous system that evolved to react to those objects in adaptive ways” (p. 192).

Modularity is controversial in cognitive science. Most psychologists accept Fodor’s (1983) claim that many aspects of perceptual and linguistic processing are the output of modules, which are informationally encapsu-
lated special purpose processing mechanisms. Informational encapsulation means that the module works on its own proprietary inputs. Knowledge contained elsewhere in the mind will not affect the output of the module. For example, knowing that two lines are the same length in the Muller-Lyer illusion does not alter the percept that one line is longer. However, Fodor himself rejects the idea that much of higher cognition can be understood as the output of modules. On the other hand, Dan Sperber (1994) has pointed out that modules for higher cognition do not need to be as tightly modularized as Fodor’s perceptual modules. All we need to say is that higher cognitive processes are modularized “to some interesting degree,” that is, higher cognition is not one big domain-general cognitive work space. There can be many bits of mental processing that are to some degree module-like. For example, quick, strong, and automatic rejection of anything that seems like incest suggests the output of an anti-incest module, or modular intuition. (See the work of Debra Lieberman, volume 1 of this collection.) Even when the experimenter explains that the brother and sister used two forms of birth control and that the sister was adopted into the family at age 14, many people still say they have a gut feeling that it is wrong for the siblings to have consensual sex. The output of the module is not fully revisable by other knowledge, even though some people overrule their intuition and say, uneasily, that consensual adult sibling incest is OK.

We do not know what point on the continuum from simple preparedness to hard and discrete modularity is right, so we tentatively adopt Sperber’s intermediate position that there are a great many bits of mental processing that are modular “to some interesting degree.” (We see no reason to privilege the blank-slate side of the continuum as the default or “conservative” side.) Each of our five foundations can be thought of either as a module itself or, more likely, as a “learning module”—a module that generates a multiplicity of specific modules during development within a cultural context (e.g., as a child learns to recognize in an automatic and module-like way specific kinds of unfairness, or of disrespect; see Haidt & Joseph, in press, and Sperber, 2005, for details). We particularly like Sperber’s point that “because cognitive modules are each the result of a different phylogenetic history, there is no reason to expect them all to be built on the same general pattern and elegantly interconnected” (Sperber, 1994, p. 46). We are card-carrying antiparsimonists. We believe that psychological theories should have the optimum amount of complexity, not the minimum that a theorist can get away with. The history of moral psychology is full of failed attempts to derive all of morality from a single
Question 5: How Does Morality Develop?

Once you see morality as grounded in a set of innate moral modules (Sperber modules, not Fodor modules), the next step is to explain how children develop the morality that is particular to their culture and the morality that is particular to themselves. The first of two main tools we need for an intuitionist theory of development is “assisted externalization” (see Fiske, 1991). The basic idea is that morality, like sexuality or language, is better described as emerging from the child (externalized) on a particular developmental schedule rather than being placed into the child from outside (internalized) on society’s schedule. However, as with linguistic and sexual development, morality requires guidance and examples from the local culture to externalize and configure itself properly, and children actively seek out role models to guide their development. Each of the five foundations matures and gets built upon at a different point in development—for example, 2-year-olds are sensitive to suffering in people and animals (Zahn-Waxler & Radke-Yarrow, 1982), but they show few concerns for fairness and equal division of resources until some time after the third birthday (Haidt, Lobue, Chiong, Nishida, & Deloachc.e, in preparation), and they do not have a full understanding of purity and contagion until around the age of 7 or 8 (Rozin, Fallon, & Augustoni-Ziskind, 1986). When their minds are ready, children will begin showing concerns about and emotional reactions to various patterns in their social world (e.g., suffering, injustice, moral contamination). These reactions will likely be crude and inappropriate at first, until they learn the application rules for their culture (e.g., share evenly with siblings, but not parents) and until they develop the wisdom and expertise to know how to resolve conflicts among intuitions.

Take, for example, the game of cooties (Haidt et al., in preparation). All over the United States children play a game from roughly ages 8 to 10 in which some children are said to have “cooties,” which are a kind of invisible social germ. Cooties reflects three principal concerns: sex segregation (boys think girls have cooties, and vice versa), social popularity (children
who are unattractive and of low social status are much more likely to have cooties), and hygiene (children who are physically dirty are more likely to have cooties). Cooties are spread by physical contact, and they are eliminated by receiving a symbolic “cooties shot,” making it clear that cooties relies heavily on children’s intuitions about purity, germs, and disease. Cooties is not supported by society at large or by the media—in fact, adults actively oppose cooties, because the game is often cruel and exclusionary. One might still say that cooties is simply learned from other children and passed on as part of peer culture, the way that Piaget (1965/1932) showed that the game of marbles is passed on. And this is certainly correct. But one must still ask: why do some games persist for decades or centuries while other games (e.g., educational games made up by adults) do not get transmitted at all? Cooties, for example, is found in some form in many widely separated cultures (Hirschfeld, 2002; Opie & Opie, 1969).

The game of cooties is so persistent, stable, and ubiquitous, we believe, because it is a product of the maturation and elaboration of the purity foundation. When children acquire the cognitive ability of contamination sensitivity around the age of 7, they begin applying it to their social world. Suddenly, children who are disliked, and the opposite sex in general, come to be felt to be contaminating—their very touch will infect a person with their dislikable essence. Children’s culture is creative, and children mix in other elements of their experience, such as getting vaccines to prevent disease. However, the critical point here is that the cooties game would not exist or get transmitted if not for the purity foundation; the game is both enabled and constrained by the structure of children’s minds and emotions. The game is a product of assisted externalization as each cohort of children teaches the game to the next, but only when their minds are ready to hold it (see Sperber & Hirschfeld, 2004, on the cognitive foundations of cultural transmission).

The second crucial tool for an intuitionist theory of moral development is a notion of “virtues as constrained social constructions.” Virtues are attributes of a person that are to some degree learned or acquired. Philosophers since Aristotle have stressed the importance of habit and practice for the development of virtues, and parents, schools, and religious organizations devote a great deal of effort to the cultivation of virtues in young people. The philosopher Paul Churchland offers an approach to virtue tailored for modern cognitive science. He sees virtues as skills a child develops that help her navigate the complex social world. Virtues are “skills of social perception, social reflection, imagination, and reasoning, and social navigation and manipulation that normal social learning produces”
Virtues, as sets of culturally ideal skills, clearly vary around the world and across cultures. Even within a single culture, the virtues most highly valued can change over the course of a single generation, as happened in the some parts of the Western world with the so-called “generation gap” of the 1960s. Yet virtues, like gods and ghosts, do not vary wildly or randomly (Boyer, 2001). Lists of focal virtues from around the world usually show a great deal of overlap (Peterson & Seligman, 2004). Virtue theorists such as Churchland are often silent on the issue of constraint, suggesting implicitly that whatever virtues a society preaches and reinforces will be the ones that children develop. Yet such a suggestion is an endorsement of equipotentiality, which has been thoroughly discredited in psychology. There is no reason to suppose that every virtue is equally learnable. Virtue theories can be greatly improved—not vitiated—by adding in a theory of constraint. The constraints we suggest are the five foundations of intuitive ethics.

Some virtues seem to get constructed on a single foundation. For example, as long as people have intuitions about harm and suffering, anyone who acts to relieve harm and suffering will trigger feelings of approval. The virtue of kindness is a social construction that a great many cultures have created to recognize, talk about, and reward people who act to relieve suffering. What it means to be kind will vary to some degree across cultures, but there will be a family resemblance among the exemplars. A similar story can be told for virtues such as honesty (for the fairness foundation), self-sacrifice (in-group), respect (authority), and cleanliness (purity). However, other virtues are much more complex. Honor, for example, is built upon the authority foundation in most traditional cultures (honor is about the proper handling of the responsibilities of high rank), as well as upon fairness/reciprocity (an honorable man pays his debts and avenges insults) and purity (honor is pure and cannot tolerate any stain). But honor is often quite different for women (drawing more heavily on the virtue of chastity, based on the purity foundation; see Abu-Lughod, 1986), and particular notions of honor vary in dramatic yet predictable ways along with the social and economic structure of any given society (e.g., herding vs. agricultural cultures; Nisbett & Cohen, 1996). (For more on foundations, modules, and virtues, see Haidt & Joseph, in press.)

Moral development can now be understood as a process in which the externalization of five (or more) innate moral modules meets up with a
particular set of socially constructed virtues. There is almost always a close match, because no culture can construct virtues that do not mesh with one or more of the foundations. (To do so is to guarantee that the next generation will alter things, as they do when converting a pidgin language to a Creole.) Adults assist the externalization of morality by socializing for virtue, but they often overestimate their causal influence because they do not recognize the degree to which they are going with the flow of the child’s natural moral proclivities. Adults may also overestimate their influence because children from middle childhood through adolescence are looking more to their peers for moral attunement than to their parents (Harris, 1995). The social parts of the SIM call attention to the ways that moral judgments made by children, especially high-status children, will spread through peer networks and assist in the externalization of intuitions and the construction of virtues.

The five foundations greatly underspecify the particular form of the virtues and the constellation of virtues that will be most valued. As with cuisine, human moralities are highly variable, but only within the constraints of the evolved mind. One of the most interesting cultural differences is the current “culture war” between liberals and conservatives in the United States and in some other Western cultures. The culture war can be easily analyzed as a split over the legitimacy of the last three foundations (Haidt & Graham, in press-a, in press-b). All cultures have virtues and concerns related to harm/care and fairness/reciprocity. However cultures are quite variable in the degree to which they construct virtues on top of the in-group/loyalty, authority/respect, and purity/sanctity foundations. American liberals in particular seem quite uncomfortable with the virtues and institutions built on these foundations, because they often lead to jingoistic patriotism (in-group), legitimization of inequality (authority), and rules or practices that treat certain ethnic groups as contagious (purity, as in the segregation laws of the American South). Liberals value tolerance and diversity and generally want moral regulation limited to rules that protect individuals, particularly the poor and vulnerable, and that safeguard justice, fairness, and equal rights. Conservatives, on the other hand, want a thicker moral world in which many aspects of behavior, including interpersonal relations, sexual relations, and life-or-death decisions are subject to rules that go beyond direct harm and legal rights. Liberals are horrified by what they see as a repressive, hierarchical theocracy that conservatives want to impose on them. Conservatives are horrified by what they see as the “anything goes” moral chaos that liberals have created, which many see as a violation of the will of God and as a threat to their
efforts to instill virtues in their children (Haidt, 2006, chapter 9; Haidt & Graham, in press-a).

**Question 6: Why do People Vary in Morality?**

If virtues are learned skills of social perception, reflection, and behavior, then the main question for an intuitionist approach to moral personality is to explain why people vary in their virtues. The beginning of the story must surely be innate temperament. The “first law of behavioral genetics” states that “all human behavioral traits are heritable” (Turkheimer, 2000, 160). On just about everything ever measured, from liking for jazz and spicy food to religiosity and political attitudes, monozygotic twins are more similar than are dizygotic twins, and monozygotic twins reared apart are usually almost as similar as those reared together (Bouchard, 2004). Personality traits related to the five foundations, such as disgust sensitivity (Haidt, McCauley, & Rozin, 1994) or social dominance orientation (which measures liking for hierarchy versus equality; Pratto, Sidanius, Stallworth, & Malle, 1994), are unlikely to be magically free of heritability. The “Big Five” trait that is most closely related to politics—openness to experience, on which liberals are high—is also the most highly heritable of the five traits (McCrae, 1996). Almost all personality traits show a frequency distribution that approximates a bell curve, and some people are simply born with brains that are prone to experience stronger intuitions from individual moral modules (Link 1 in figure 4.1).

Learning, practice, and the assistance of adults, peers, and the media then produce a “tuning up” as each child develops the skill set that is her unique pattern of virtues. This tuning up process may lead to further building upon (or weakening of) particular foundations. Alternatively, individual development might be better described as a broadening or narrowing of the domain of application of a particular module. A moralist is a person who applies moral intuitions and rules much more widely than do other members of his culture, such that moral judgments are produced by seemingly irrelevant cues.

A major source of individual differences may be that all children are not equally “tunable.” Some children are more responsive to reward and punishment than others (Kochanska, 1997). Some people tend to use preexisting internal mechanisms for quick interpretation of new information; others have more conservative thresholds and gather more information before coming to a judgment (Lewicki, Czyzewska, & Hill, 1997). Children
who are less responsive to reward and who do more thinking for themselves can be modeled as being relatively less influenced by the social persuasion link (link 4 in figure 4.1). They may be slower to develop morally, or they may be more independent and less conventional in their final set of virtues. Individual differences in traits related to reasoning ability, such as IQ or need for cognition (Cacioppo & Petty, 1982), would likely make some people better at finding post hoc arguments for their intuitions (link 2) and at persuading other people via reasoned argument (link 4). Such high-cognition people might also be more responsive themselves to reasoned argument and also better able to engage in reasoning that contradicts their own initial intuitions (links 5 and 6).

A big question in moral personality is the question of behavior: Why do some people act ethically and others less so? Much of modern social psychology is a warning that the causes of behavior should not be sought primarily in the dispositions (or virtues) of individuals (Ross & Nisbett, 1991). John Doris (2002) has even argued that the underappreciated power of situations is a fatal blow for virtue theories. However, we believe such concerns are greatly overstated. The only conception of virtue ruled out by modern social psychology is one in which virtues are global tendencies to act in certain ways (e.g., courageous, kind, chaste) regardless of context. That is the position that Walter Mischel (1968) effectively demolished, showing instead that people are consistent across time within specific settings. Our conception of virtue as a set of skills needed to navigate the social world explicitly includes a sensitivity to context as part of the skill. One reason it takes so long to develop virtues is that they are not simple rules for global behavior. They are finely tuned automatic (intuitive) reactions to complex social situations. They are a kind of expertise.

However, even with that said, it is still striking that people so often fail to act in accordance with virtues that they believe they have. The SIM can easily explain such failures. Recall the Robert Wright quote that the brain is “a machine for winning arguments.” People are extraordinarily good at finding reasons to do the things they want to do, for nonmoral reasons, and then mounting a public relations campaign to justify their actions in moral terms. Kurzban and Aktopis (2006) recently surveyed the literature on self-presentation to argue that the modularity of the human mind allows people to tolerate massive inconsistencies between their private beliefs, public statements, and overt behaviors. Hypocrisy is an inevitable outcome of human mental architecture, as is the blindness to one’s own hypocrisy.
Unresolved Questions

The SIM is a new theory, though it has very old roots. It seems to handle many aspects of morality quite easily; however, it has not yet been proven to be the correct or most powerful theory of morality. Many questions remain; much new evidence is needed before a final verdict can be rendered in the debate between empiricist, rationalist, and moral sense theories. Here we list some of those questions.

1. What is the ecological distribution of types of moral judgment? The SIM claims that people make the great majority of their moral judgments using quick intuitive responses but that sometimes, not often, people reject their initial intuition after a process of “private reflection” (or, even more rarely, directly reasoned judgment). Rationalist theorists claim that true private moral reasoning is common (Pizarro & Bloom, 2003). An experience sampling or diary study of moral judgment in daily life would be helpful in settling the issue. If moral reasoning in search of truth (with conscious examination of at least one reason on both sides, even when the person has an interest in the outcome) were found to happen in most people on a daily basis, the SIM would need to be altered. If, on the other hand, moral judgments occur several or dozens of times a day for most people, with “cognitive conflicts” or overridden intuitions occurring in less than, say, 5% of all judgments, then the SIM would be correct in saying that private moral reasoning is possible but rare.

2. What are the causes of moral persuasion and change? The SIM posits two links—the reasoned persuasion link and the social persuasion link. These links are similar to the central and peripheral processes of Petty and Cacioppo’s (1986) elaboration-likelihood model of persuasion, so a great deal of extant research can be applied directly to the moral domain. However there are reasons to think that persuasion may work differently for moral issues. Skitka (2002) has shown that when people have a “moral mandate”—when they think they are defending an important moral issue—they behave differently, and are more willing to justify improper behavior, than when they have no moral mandate. Further research is needed on moral persuasion.

3. Are all intuitions externalized? We believe the most important ones are, but it is possible that some intuitions are just moral principles that were once learned consciously and now have become automatic. There is no innate knowledge of shoe-tying, but after a few thousand times the act becomes automatic and even somewhat hidden from conscious introspection. Might some moral principles be the same? We do not believe that
the intuitions we have talked about can be explained in this way—after all, were you ever explicitly told not to have sex with your siblings? But perhaps some can. Is it possible to create a moral intuition from scratch which does not rely on any of the five intuitive foundations and then get people to really care about it? (An example might be the judgment we have seen in some academic circles that groups, teams, or clubs that happen to be ethnically homogeneous are bad, while ethnic diversity is, in and of itself, good.)

4. Is there a sensitive period for learning moral virtues? Haidt (2001) suggested that the period when the frontal cortex is myelinating, from roughly ages 7 through 15 or so, might be a sensitive period when a culture’s morality is most easily learned. At present there is only one study available on this question (Minoura, 1992).

5. Can people improve their moral reasoning? And if they did, would it matter? It is undoubtedly true that children can be taught to think better about any domain in which they are given new tools and months of practice using those tools. However, programs that teach thinking usually find little or no transfer outside of the classroom (Nickerson, 1994). And even if transfer were found for some thinking skills, the SIM predicts that such improvements would wither away when faced with self-interest and strong gut feelings. (It might even be the case that improved reasoning skills improve people’s ability to justify whatever they want to do.) A good test of rationalist models versus the SIM would be to design a character education program in which one group receives training in moral reasoning, and the other receives emotional experiences that tune up moral sensitivity and intuition, with guidance from a teacher or older student. Which program would have a greater impact on subsequent behavior?

**Philosophical Implications**

The SIM draws heavily on the work of philosophers (Hume, Gibbard, Aristotle), and we think it can give back to philosophy as well. There is an increasing recognition among philosophers that there is no firewall between philosophy and psychology and that philosophical work is often improved when it is based on psychologically realistic assumptions (Flanagan, 1991). The SIM is intended to be a statement of the most important facts about moral psychology. Here we list six implications that this model may have for moral philosophy.

1. **Moral truths are anthropocentric truths** On the story we have told, all cultures create virtues constrained by the five foundations of intuitive ethics.
Moral facts are evaluated with respect to the virtues based in these underlying intuitions. When people make moral claims they are pointing to moral facts outside of themselves—they intend to say that an act is in fact wrong, not just that they disapprove of it. If there is a nontrivial sense in which acts are in fact wrong, then subjectivist theories are wrong too. On our account, moral facts exist, but not as objective facts which would be true for any rational creature anywhere in the universe. Moral facts are facts only with respect to a community of human beings that have created them, a community of creatures that share a “particular fabric and constitution,” as Hume said. We believe that moral truths are what David Wiggins (1987a) calls “anthropocentric truths,” for they are true only with respect to the kinds of creatures that human beings happen to be. Judgments about morality have the same status as judgments about humor, beauty, and good writing. Some people really are funnier, more beautiful, and more talented than others, and we expect to find some agreement within our culture, or at least our parish, on such judgments. We expect less agreement (but still more than chance) with people in other cultures, who have a slightly different fabric and constitution. We would expect intelligent creatures from another planet to show little agreement with us on questions of humor, beauty, good writing, or morality. (However, to the extent that their evolutionary history was similar to our own, including processes of kin selection and reciprocal altruism, we would expect to find at least some similarity on some moral values and intuitions.)

2. The naturalistic imperative: All ought statements must be grounded, eventually, in an is statement  If moral facts are anthropocentric facts, then it follows that normative ethics cannot be done in a vacuum, applicable to any rational creature anywhere in the universe. All ethical statements should be marked with an asterisk, and the asterisk refers down to a statement of the speaker’s implicit understanding of human nature as it is developed within his culture. Of course, the kind of is-to-ought statements that Hume and Moore warned against are still problematic (e.g., “men are bigger than women so men ought to rule women”). But there is another class of is-to-ought statements that works, for example, “Sheila is the mother of a 4-year-old boy, so Sheila ought*3 to keep her guns out of his reach.” This conclusion does not follow logically from its premise, but it is instantly understood (and probably endorsed) by any human being who is in full possession of the anthropocentric moral facts of our culture. If Greene (volume 3 of this collection) is correct in his analysis of the psychological origins of deontological statements, then even meta-ethical work must be marked with an asterisk, referring down to a particular understanding of
human nature and moral psychology. When not properly grounded, entire schools of meta-ethics can be invalidated by empirical discoveries, as Greene may have done.

3. **Monistic theories are likely to be wrong**  If there are many independent sources of moral value (i.e., the five foundations), then moral theories that value only one source and set to zero all others are likely to produce psychologically unrealistic systems that most people will reject. Traditional utilitarianism, for example, does an admirable job of maximizing moral goods derived from the harm/care foundation. However, it often runs afoul of moral goods derived from the fairness/reciprocity foundation (e.g., rights), to say nothing of its violations of the in-group/loyalty foundation (why treat outsiders equal to insiders?), the authority/respect foundation (it respects no tradition or authority that demands anti-utilitarian practices), and the purity/sanctity foundation (spiritual pollution is discounted as superstition). A Kantian or Rawlsian approach might do an admirable job of developing intuitions about fairness and justice, but each would violate many other virtues and ignore many other moral goods. An adequate normative ethical theory should be pluralistic, even if that introduces endless difficulties in reconciling conflicting sources of value. (Remember, we are antiparsimonists. We do not believe there is any particular honor in creating a one-principle moral system.) Of course, a broad enough consequentialism can acknowledge the plurality of sources of value within a particular culture and then set about maximizing the total. Our approach may be useful to such consequentialists, who generally seem to focus on goods derived from the first two foundations only (i.e., the “liberal” foundations of harm and reciprocity).

4. **Relativistic and skeptical theories go too far**  Meta-ethical moral relativists say that “there are no objectively sound procedures for justifying one moral code or one set of moral judgments as against another” (Neilsen, 1967, 125). If relativism is taken as a claim that no one code can be proven superior to all others, then it is correct, for given the variation in human minds and cultures, there can be no one moral code that is right for all people, places, and times. A good moral theory should therefore be pluralistic in a second sense in stating that there are multiple valid moral systems (Shweder & Haidt, 1993; Shweder et al., 1997). Relativists and skeptics often go further, however, and say that no one code can be judged superior to any other code, but we think this is wrong. If moral truths are anthropocentric truths, then moral systems can be judged on the degree to which they violate important moral truths held by members of that society. For example, the moral system of Southern White slaveholders
radically violated the values and wants of a large proportion of the people involved. The system was imposed by force, against the victims’ will. In contrast, many Muslim societies place women in roles that outrage some egalitarian Westerners, but that the great majority within the culture—including the majority of women—endorse. A well-formed moral system is one that is endorsed by the great majority of its members, even those who appear, from the outside, to be its victims. An additional test would be to see how robust the endorsement is. If Muslim women quickly reject their society when they learn of alternatives, the system is not well formed. If they pity women in America or think that American ways are immoral, then their system is robust against the presentation of alternatives.

5. The methods of philosophical inquiry may be tainted If the SIM is right and moral reasoning is usually post hoc rationalization, then moral philosophers who think they are reasoning their way impartially to conclusions may often be incorrect. Even if philosophers are better than most people at reasoning, a moment’s reflection by practicing philosophers should bring to mind many cases where another philosopher was clearly motivated to reach a conclusion and was just being clever in making up reasons to support her already-made-up mind. A further moment of reflection should point out the hypocrisy in assuming that it is only other philosophers who do this, not oneself. The practice of moral philosophy may be improved by an explicit acknowledgment of the difficulties and biases involved in moral reasoning. As Greene (volume 3 of this collection) has shown, flashes of emotion followed by post hoc reasoning about rights may be the unrecognized basis of deontological approaches to moral philosophy.

Conclusion

When the SIM was first published (Haidt, 2001), some people thought the model had threatening implications for human dignity. They thought the model implied that people are dumb and morality is fixed by genes, so there is no possibility of moral progress. The model does state that moral reasoning is less trustworthy than many people think, so reasoning is not a firm enough foundation upon which to ground a theory—normative or descriptive—of human morality. However, the alternative to reason is not chaos, it is intuition. Intuitive and automatic processes are much smarter than many people realize (Bargh & Chartrand, 1999; Gladwell, 2005). Intuitions guide the development of culture-specific virtues. A fully enculturated person is a virtuous person. A virtuous person really cares about things...
that happen in the world, even when they do not affect her directly, and she will sometimes take action, even when it does not seem rational to do so, to make the world a better place. We believe that social intuitionism offers a portrait of human morality that is just as flattering as that offered by rationalism, yet much more true to life.

Notes

We thank Cate, Birtley, Josh Greene, Ben Shear, and Walter Sinnott-Armstrong for helpful comments on earlier drafts.

1. Haidt (2001) had defined moral intuition as “the sudden appearance in consciousness of a moral judgment,” (p. 818) thereby conflating the intuition, the judgment, and Link 1 into a single psychological event, obviating any need for the link. We thank Walter Sinnott-Armstrong for pointing out this error and its solution.

2. Haidt and Joseph (2004) talked about only the first four moral modules, referring to the in-group module only in a footnote stating that there were likely to be many more than four moral modules. In a subsequent publication (Haidt & Joseph, in press) we realized that in-group concerns were not just a branch of authority concerns and had to be considered equivalent to the first four.

3. It is an anthropocentric fact that motherhood requires loving and caring for one’s children. There could be intelligent species for whom this is not true.