

**Social Intuitionists Answer Six Questions about Moral Psychology**

Jonathan Haidt, University of Virginia (USA)

Fredrik Bjorklund, Lund University (Sweden)

February 14, 2005

Draft 1.0, Comments Welcome

To appear in: W. Sinnott-Armstrong (Ed.), *Moral Psychology*, Vol. 3

[Word count: 13,960 for main text; 16,380 for whole document]

Contact information:

Jonathan Haidt, [Haidt@virginia.edu](mailto:Haidt@virginia.edu);

Fredrik Bjorklund, [Fredrik.Bjorklund@psychology.lu.se](mailto:Fredrik.Bjorklund@psychology.lu.se)

Here are two of the biggest questions in moral psychology: 1) Where does morality 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) Morality comes from a small set of evolved intuitions built into the brain by evolution, which constrain and enable 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. But there is more: moral reasoning within a single person's head is usually devoted to finding reasons to support one'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." 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, 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 Saltzstein & Kasachkoff, 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 social intuitionist model. We follow up with question 3: what is the evidence for the social intuitionist model? We then address three questions that we believe become answerable in a coherent and consistent way via the social intuitionist model. 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 no 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 does morality 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 assumed that, as with the magician's rabbit, morality had to 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 20<sup>th</sup> 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 are exposed to 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 alternatives to moral empiricism are moral nativism – the idea that the mind contains some innate moral knowledge or moral faculty – and moral rationalism – the idea that what is innate is rationality, which a child uses to reach an increasingly adequate or accurate

moral understanding. Rationalist approaches to morality usually posit relatively little specific content – perhaps a few a priori rules or concepts. The emphasis instead is on the act of construction, on the way that a child builds up her own moral understanding 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 is determined by what gets rewarded and punished. But as their cognitive abilities mature around the ages of 6-8 and they become able to "de-center," 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. However cognitive development is just a pre-requisite 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 a passive receptacle for social conditioning. Eliot 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 three domains of social knowledge for themselves.

To give you a sense of a rationalist approach we report the transcript of a remarkable interview that one of us (JH) 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 four 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]

Int: *Dad, what would happen if I pooped in the sink?*

Sub: *The people who work here would get mad at you.*

Int: *What would happen if I pooped in the sink at home?*

Sub: *I'd get mad at you.*

Int: *What would happen if YOU pooped in the sink at home?*

Sub: *Mom would get mad at me.*

Int: *Well, what would happen if we ALL pooped in the sink at home?*

Sub: [pause...] *I guess we'd all get in trouble.*

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

Sub: *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 Kohlbergian stage: things seem to be wrong because they are punished. But note the skill and persistence of the interviewer, who probes for a deeper answer by changing the transgression to remove a punishing agent. However 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 imminent 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 are watchdogs of the moral world (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 nativism – of which moral sense theories are the most widely discussed.

When God began to recede from scientific explanations in the 16<sup>th</sup> century, some philosophers began to wonder if God was really needed to explain morality either. In the 17<sup>th</sup> and 18<sup>th</sup> 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 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*. (Enquiry Concerning the Principles of Morals, 1777/1960, 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 does morality come from? It comes 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. But this statement is also extraordinary as a statement about the controversy "started of late." Hume's statement is just as true in 2005 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; J. Q. Wilson, 1993; Haidt, 2001). We will not try to be fair and unbiased guides through this debate (indeed, our theory says you should not trust 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 innate 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.

**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 human beings are 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 unknown 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 social intuitionist model [SIM], presented in Figure 1 (from Haidt, 2001). The model is composed of 6 links, each of which describes a psychological process. The first four links are the core of the model, intended to capture the great majority of judgments for most people. Links 5 and 6 are hypothesized to occur rarely, but should be of great interest to philosophers because philosophers use these links far more than most people (Kuhn, 1991). The existence of each link or process is well supported by psychological research, presented below. However whether everyday moral judgment is best captured by this particular arrangement of processes is still controversial (Pizarro & Bloom, 2003), so the SIM should be considered a hypothesis for now, rather than an established fact. The 6 links are as follows:

*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 judgments are made quickly, effortlessly, and intuitively. We see an act of violence, 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. This judgment is Link 1, the intuitive judgment link, in Figure 1. Moral intuitions are defined as: *the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion* (Haidt, 2001, p.818; for more on intuition, see question 4 below). This is the “finer internal sense” that Hume talked about.

These flashes of intuition are not dumb; as with the superb software that runs visual perception, they often hide a great deal of sophisticated 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 perceptual 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. (Kahneman, 2004, p.26)



However even if moral judgments are made intuitively, we often feel a pressure 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 intuitive 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 a few 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, 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 (for example, 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 coke.” 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) 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 moral 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 people track the reputations of other people and manipulate people to enhance one’s own reputation (Dunbar, 1996). People are able to re-use this tool for new purposes, including scientific or philosophical reasoning, 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 social intuitionist model. 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 and Haidt, in prep.). 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 prep.; Wright 1994). Allan Gibbard (1990) is perhaps the philosopher who was 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. But 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, p.280) 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.

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 not on 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 includes 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). One might say that these two sentences contain 7 arguments against altering female genitalia, each indicated in italics, but we think this passage is more accurately described as an attempt to push 7 different buttons, triggering seven different flashes of intuition in the listener. Rhetoric is the art of pushing the ever-evaluating mind over to the side you want 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 in 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. Over and over again, people adjust their beliefs to fit with the beliefs of others, not just because they assume others have useful information, but because 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 mammal is the naked mole rat 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. Human beings alone cooperate widely and intensely with non-kin, 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 social intuitionist model. 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 only 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 true moral reasoning in private? Yes. The loop described by the first 4 links in the SIM is intended to capture the great majority of moral judgments made by the great majority of people. But 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. 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.

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 consistently humane in striving to minimize the suffering of sentient beings, but it leads to the conclusion that the life of a healthy chimpanzee deserves greater protection than that of an acephalic infant human that 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 as a paragon of evil. (See also Derek Parfitt’s [1984] conclusion that we should populate the world much more fully, and Kant’s (1785/1959) 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 oneself into the shoes of another person one 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 allowing reason to choose among the alternatives based on the conscious application of a rule or principle. This pathway amounts to having an inner dialogue with oneself (Tappan, 1997), obviating the need for a discourse partner. Is this really reasoning? Yes, as long as one allows that flashes of intuition play an important role in reasoning. William James captured 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)

People sometimes do look on both sides of an issue, thereby triggering intuitions on both sides. Sometimes reasoning may truly adjudicate such affective deadlocks; other times one intuition is stronger and it sends reasoning off to look for more supporting evidence. But it must be stressed that such deadlocks are fairly rare in our moral lives, and they can only happen where there is no strong emotion or prior commitment operating. If you don't believe this, try this challenge: 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 hurt your head to examine those reasons and weigh the pros and cons dispassionately. 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, then you are likely to be either a psychopath or a philosopher. Philosophers are one of the only groups that have been found to spontaneously look for reasons on both sides of a question (Kuhn, 1991). It seems plausible, furthermore, that people who choose to go in to philosophy are people who are unusually capable of thinking about 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 reasoning and intuition. What is the evidence for these claims? In this section we briefly summarize the findings from relevant empirical studies.

In the 1980s a debate arose between Eliot Turiel (1983), who said that the moral domain is universally limited to issues of harm, rights and justice, and Richard Shweder (Shweder et al.,

1987), who said that the moral domain is variable across cultures. Shweder, Mahapatra and Miller (1987) showed that in Orissa, India, the moral domain expands to include a great many issues related to food, clothing, sex roles, and other practices Turiel would label as social conventions. However Turiel, Killen, & Helwig (1987) argued that most of Shweder's research vignettes contained harm, once you understand how Indians understood 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). Subjects were adults or children (ages 10-12; they did not receive the chicken story); they were of high or of low social class; and they were residents of 3 cities: Recife, Brazil, Porto Alegre, Brazil, and Philadelphia, USA. The basic finding was that the high social class adult groups, which were all 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 behavior would not be wrong in another culture where it was 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 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 et al. (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).

However while conducting these interviews, 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, participants rarely changed their minds. Rather, they searched harder for additional reasons, sometimes laughing and confessing that they could not explain themselves. Haidt & 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 and Murphy, 2005) brought moral dumbfounding into the lab to examine it more closely. In study 1 we gave subjects 5 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). But 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 participants to a cognitive load – an attention task that took up some of their conscious mental workspace – and found that this load increased the level of moral dumbfounding without changing subjects' judgments or their level of persuadability .

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.” Participants then read 6 moral judgment stories each of which included either the word “take” or the word “often.” Only highly hypnotizable subjects who were amnesic for the post-hypnotic suggestion were used. In two studies, the flash of disgust that participants felt while reading 3 of their 6 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 phenomenon: ““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. But one third of all subjects who had 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, and Clore (2005) 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.

A great deal of neuroscience research supports the idea that flashes of affect are essential for moral judgment. 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. 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 have studied the brains of healthy people making moral judgments while in an fMRI scanner (Greene et al., 2001). They found that the distinctions people make between various moral dilemmas are well-predicted by whether or not certain brain areas involved in emotional responding light up. When considering pushing one man off of a bridge 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. When they think about throwing a switch to shift the train from killing five to killing one they have no such flash, and they choose the utilitarian response.

Greene’s data does show that people sometimes overrule their gut feelings and choose the utilitarian response in the Bridge dilemma – in fact, Greene can tell who is likely to do so by observing who shows a slow surge of activity in the dorsolateral prefrontal cortex, an area known to support conscious reasoning. So Greene believes that the SIM underestimates the ability of reasoning to contradict intuition. However at the moment we cannot know if such over-rulings are true cases of reason versus intuition, or of a double intuition, in which some people also think about the five men who will be killed, feel a flash of affect, and then engage in reasoning to sort out the conflict. (This is exactly the process that Link 6, the reasoned reflection link, was designed to capture). Greene may ultimately be correct that in cases such as these, where intuitions and emotions are fairly subtle, some people (particularly those who are high on the personality trait of “need for cognition” (Cacioppo & Petty, 1982) do use reasoning as more than

a slave of the passions. However we believe that in cases of strong intuitions, as is typical in real life situations where one's own interests and friendships are at stake, the stronger claims of the SIM are still the best description of human moral judgment. Further research will be needed to resolve this question.

[\*\*should I add more here? Skitka? Discuss relative lack of work on the social side of the model?]

#### **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 whatever 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 or the media. However we see two strong arguments against such empiricist approaches. The first, pointed out by Tooby, Cosmides, & Barrett (in press), 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 leader can select a group of unusual adults able to believe in universal love and oppose all forms of hatred or jealousy, nobody has ever been able to raise a 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). But 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 that underlies the enormous diversity of moral “cuisines?” Just such an analogy was made by the Chinese philosopher Mencius 2400 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 designed to capture what is universal about human cultures: Donald Brown’s (1991) catalogue “Human Universals,” and Alan Fiske’s (1992) grand integrative theory 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 5) that each element appeared in. The winners, showing up clearly in all five works, were suffering (a sensitivity to or dislike of signs

of physical pain in others), reciprocity (a set of emotional responses related to playing tit-for-tat, such as negative responses to those who fail to repay favors), and hierarchy (a set of concerns about navigating status hierarchies, for example anger towards 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 himself specifically included suffering (commiseration) and hierarchy (shame, respect, and reverence) as human universals.

We tried to see how much moral work these three sets of intuitions could do, but quickly realized we would need to bring in others that did not get 5 votes. The two that were most needed were concerns about purity (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 ingroup and outgroup<sup>1</sup>.

We (Haidt, Joseph, and 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 already been told many times. We hope nobody will find it controversial to suppose that evolution has built in to humans (and to some extent chimpanzees, bonobos, and other social mammals) an emotional sensitivity to issues related to suffering/caretaking, reciprocity/alliance, hierarchy/social-order, and ingroup/outgroup. The only set of intuitions with no clear precursor in other animals is purity/pollution. But concerns with 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 (P. Rozin *et al.*, 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 some 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,

---

<sup>1</sup> Haidt and Joseph (2004) talked about only the first four moral modules, referring to the ingroup module only in a footnote stating that there were likely to be many more than four moral modules. In a subsequent publication (Haidt & Joseph, 2005) we realized that ingroup concerns were so important that they had to be considered equivalent to the first four.

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. So 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 suffering, reciprocity, hierarchy, ingroups, and purity, however they 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 module 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 towards feces and rotting meat, and also towards 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, p. 192) 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.”

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 encapsulated 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, that higher cognition is not one big domain-general cognitive workspace. 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. 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 massive modularity is right, so we adopt Sperber's position that there are a great many bits of mental processing that are modular "to some interesting degree." Each of our five foundational intuitions can be thought of as a module, or set of functionally related modules. 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 anti-parsimonists. We believe that psychological theories should have the optimum amount of complexity, not the minimum. The history of moral psychology is full of failed attempts to derive all of morality from a single source (e.g., non-contradiction, or empathy, or internalization). We think it makes much more sense to look at morality as a set of multiple concerns about social life, each one with its own history and mechanism.

### **Question 5: How does morality develop?**

Once you see morality as grounded in a set of innate moral modules, the next step is to explain how children develop the morality that is particular to their culture, and to themselves. The first of two main tools we need for an intuitionist theory of development is *assisted externalization* (see Fiske, 1991). Each of the five moral modules matures at a different point in development – for example, two-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 around the age of 4, and they do not have a full understanding of purity and contagion until around the age of 7 (Rozin *et al.*, 1986). As with so many other aspects of development (language, sexuality, humor...) children will begin showing abilities and



concerns at a certain point, but the expression of these new abilities is guided, constrained, suppressed or amplified by the surrounding culture. In other words, children do not just internalize whatever norms are around them; they externalize the intuitions maturing inside them, but they apply them with help from their local culture.

Take, for example, the game of cooties (Haidt *et al.*, in preparation). All over the United States children play a game from roughly ages 8-10 in which some children are said to have “cooties”, which are a kind of invisible social germ. Cooties reflects three principle 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 (for example, educational games made up by adults) do not get transmitted at all? Cooties, for example, is found in somewhat similar forms in many widely separated human 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 of the purity module. 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 were simply disliked before, including the opposite sex in general, come to be felt to be contaminating – their very touch will infect a person with their dislikeable essence. Children’s culture is creative, and children mix in other elements of their experience, such as getting vaccines to prevent disease. But the critical point here is that the cooties game would not exist or get transmitted if not for the purity module; 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.

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 at least 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” (Churchland, 1998, p.88). Moral character is then “the individual profile of [a person’s] perceptual, reflective, and behavioral skills in the social domain” (Churchland, 1998, p.89).

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 and 1970s. Yet virtues, like gods and ghosts, do not really 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 Aristotle and Churchland are usually silent on the issue of constraint, often suggesting explicitly or 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 moral modules.

Some virtues seem to grow neatly and simply out of a single module. 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 fairness (for the reciprocity module), self-sacrifice (in-group), respect (hierarchy) and cleanliness (purity). But other virtues are much more complex. Honor, for example, may draw very heavily

upon the hierarchy module in most traditional cultures (honor is about the proper handling of the responsibilities of high rank), combined with elements of reciprocity (an honorable man pays his debts and avenges all attacks) 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 in the purity module; 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 versus agricultural cultures, Nisbett & Cohen, 1996). (For more on 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 culturally constructed virtues. There is almost always a close match, because no culture can construct virtues that do not have a reasonable fit with the modules. (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 particularly focused on their peers as objects of moral attunement, rather than their parents (Harris, 1995). The social parts of the Social Intuitionist Model 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 modules greatly underspecify the particular form of the virtues and the constellation of virtues that is most highly 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 modules (Haidt & Joseph, 2004). All cultures have virtues and concerns related to harm/suffering and reciprocity/fairness. However cultures are quite variable in the degree to which they construct virtues on top of the ingroup, hierarchy, and purity modules. American liberals in particular seem quite uncomfortable with the outputs of these modules, because they often lead to jingoistic patriotism (ingroup), legitimization of inequality

(hierarchy), 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 people, 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 at what they see as a repressive, hierarchical theocracy that conservatives want to impose on them. Conservatives are horrified at 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 efforts to instill virtues in children.

### **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). 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 modules, such as disgust sensitivity (Haidt *et al.*, 1994), or social dominance orientation (which measures liking for hierarchy versus equality; Pratto *et al.* 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 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 strengthening or weakening of particular modules. Alternatively, individual development may 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 “tuneable.” 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 et al., 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 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, arguing 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 behavior. They are finely tuned automatic (intuitive) reactions to complex social situations.

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 non-moral reasons,

and then mounting a public relations campaign to justify their actions in moral terms. Kurzban and Aktipis (in press) 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 social intuitionist model 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) How often do people engage in private and two-sided moral reasoning?* The SIM claims that such reasoning is possible but rare. 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 spontaneous examination of reasons on both sides were found to happen in most people on a daily basis, the SIM would need to be altered.

*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 children to really care about it?

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. But 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 improves 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, 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 social intuitionist model 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 improved when it is based on accurate psychological facts (Flanagan,

1991). The social intuitionist model 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 facts outside of themselves, so if these facts exist in any sense then subjectivist theories are wrong. 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 (1987) 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. 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, who had an entirely different evolutionary history, to show little agreement with us on questions of morality, humor, beauty, or good writing.

2) *The naturalistic imperative: You can only derive an ought from an is.* If moral facts are anthropocentric facts then it follows that normative ethics can not 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, e.g., “Sheila *is* the mother\*<sup>2</sup> of Timmy, so Sheila *ought* to keep her guns

---

<sup>2</sup> \*It is an anthropocentric fact that motherhood requires loving and caring for one’s children.



unloaded and locked away”.

3) *Monistic theories are likely to be wrong.* If there are many independent sources of moral value (i.e., the five modules), 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.

Utilitarianism, for example, does an admirable job of maximizing moral goods derived from the suffering module. But it often runs afoul of moral goods derived from the reciprocity module (e.g., rights), to say nothing of its violations of the ingroup module (why treat outsiders equal to insiders?) the hierarchy module (it respects no tradition or authority that commands anti-utilitarian practices), and the purity module (spiritual pollution is discounted as superstition). 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.)

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, p. 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). However relativists and skeptics usually go further 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 slave holders was full of internal contradictions, which troubled even many slave holders. It was a moral system pushed into contradiction and hypocrisy by economic necessity, and critics at the time found plenty of solid ground on which to stand and deliver their criticisms.

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 sometimes 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 his 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 et al. (2001) showed, flashes of emotion may help to explain why people (including philosophers) reach different conclusions about the footbridge and trolley problems.

### **Conclusion**

When the social intuitionist model 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 that there is no possibility of moral progress. The model does state that moral reasoning is less trustworthy than many people think, so it is not a firm enough foundation upon which to ground a theory – normative or descriptive – of human morality. But the alternative to reason is not chaos, it is intuition. Intuitive and automatic processes are much smarter than many people seem to realize (Bargh & Chartrand, 1999). 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.

## References

- Abu-Lughod, L. (1986). *Veiled sentiments*. Berkeley: University of California Press.
- Asch, S. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, 70, (Whole No. 416).
- Bargh, J. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition*, 2nd edition (pp. 1-40). Hillsdale, NJ: Erlbaum.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462-479.
- Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, 126, 925-945.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497-529.
- Bjorklund, F., & Haidt, J. (in preparation). Vivid violations are morally worse. Unpublished manuscript, university of Lund.
- Bouchard, T. J. J. (2004). Genetic influence on human psychological traits: A survey. *Current Directions in Psychological Science*, 13, 148-151.
- Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York: Basic.
- Brown, D. E. (1991). *Human universals*. Philadelphia: Temple University Press.
- Burstyn, L. (1995). Female Circumcision Comes to America. *Atlantic Monthly*. October.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116-131.
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual process theories in social psychology*. New York: Guilford.
- Chan, W.-T. (1963). *A source book in Chinese philosophy*. Princeton NJ: Princeton University Press.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893-910.
- Churchland, P. M. (1998). Toward a cognitive neurobiology of the moral virtues. *Topoi*, 17, 83-96.
- Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85-116). Cambridge, UK: Cambridge University Press.
- Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Darley, J. M., & Berscheid, E. (1967). Increased liking as a result of anticipation of personal contact. *Human Relations*, 20, 29-40.
- de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.
- Doris, J. M. (2002). *Lack of character*. Cambridge, UK: Cambridge.
- Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.
- Fiske, A. P. (1991). *Structures of social life*. New York: Free Press.
- Fiske, A. P. (1992). Four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99, 689-723.
- Flanagan, O. (1991). *Varieties of moral personality: Ethics and psychological realism*. Cambridge, MA: Harvard University Press.
- Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
- Freud, S. (1976). *The interpretation of dreams* (J. Strachey, Trans. Vol. 4-5). New York: Norton.

- Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, *105*, 331-351.
- Garcia, J., & Koelling, R. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, *4*, 123-124.
- Gazzaniga, M. S. (1985). *The social brain*. New York: Basic Books.
- Gibbard, A. (1990). *Wise choices, apt feelings*. Cambridge, Mass: Harvard University Press.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fmri study of emotional engagement in moral judgment. *Science*, *293*, 2105-2108.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834.
- Haidt, J., Bjorklund, F., & Murphy, S. (2004). Moral dumbfounding: When intuition finds no reason. (unpublished manuscript, university of virginia).
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and reasons of liberals and conservatives. *Journal of Applied Social Psychology*, *31*, 191-221.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *Fall*, 55-66.
- Haidt, J., & Joseph, C. (in press). Intuitive ethics: From moral modules to cultural virtues, and back. In P. Caruthers (Ed.).
- Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613-628.
- Haidt, J., Lobue, V., Chiong, C., Nishida, T., Bloom, M., & Deloachce, J. (in preparation). Cooties and the assisted externalization of purity concerns. University of virginia.
- Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, *16*, 701-713.
- Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review*, *102*, 458-489.
- Hirschfeld, L. A. (2002). Why don't anthropologists like children?
- Hom, H., & Haidt, J. (in preparation). The bonding and norming functions of gossip. *Unpublished manuscript, University of Virginia*.
- Hume, D. (1777/1960). *An enquiry concerning the principles of morals*. La Salle, Illinois: Open Court.
- Hume, D. (1739 & 1740/1969). *A treatise of human nature*. London: Penguin.
- Kagan, J. (1984). *The nature of the child*. New York: Basic Books.
- Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3-25). New York: Russell Sage.
- Kahneman, D. (2004). Talk described in APS Observer, Vol. 17 #5, p.26 xx
- Kant, I. (1785/1959). *Foundation of the metaphysics of morals*. Indianapolis: Bobbs-Merrill.
- Kochanska, G. (1997). Multiple pathways to conscience for children with different temperaments: From toddlerhood to age 5. *Developmental Psychology*, *33*, 228-240.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Psychology and genetic epistemology* (pp. 151-235). New York: Academic Press.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674-689.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.
- Kurtines, W. M., & Gewirtz, J. L. (Eds.). (1995). *Moral development: An introduction*. Boston: Allyn and Bacon.

- Kurzban, R., & Akitipis, C. A. (in press). Modularity and the social mind: Why social psychologists should be less self-ish. In M. Schaller, J. Simpson & D. Kenrick (Eds.), *Evolution and social psychology*.
- Lapsley, D. K. (1996). *Moral psychology*. Boulder, CO: Westview.
- Latane, B., & Darley, J. M. (1970). *The unresponsive bystander*. Englewood Cliffs: Prentice Hall.
- Lewicki, P., Czyzewska, M., & Hill, T. (1997). Nonconscious information processing and personality. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 48-72). Oxford: Oxford University Press.
- McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological Bulletin*, *120*, 323-337.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371-378.
- Mineka, S., & Cook, M. (1988). Social learning and the acquisition of snake fear in monkeys. In T. R. Zentall & J. B. G. Galef (Eds.), *Social learning: Psychological and biological perspectives* (pp. 51-74). Hillsdale, N. J.: Lawrence Erlbaum.
- Minoura, Y. (1992). A sensitive period for the incorporation of a cultural meaning system: A study of Japanese children growing up in the United States. *Ethos*, *20*, 304-339.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Neilsen, K. (1967). Ethics, problems of. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 3, pp. 117-134). New York: Macmillan.
- Nickerson, R. S. (1994). The teaching of thinking and problem solving. In R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 409-449). San Diego: Academic Press.
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Boulder, CO: Westview Press.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- Opie, I., & Opie, P. (1969). *Children's games in street and playground*. Oxford, UK: Clarendon Press.
- Osgood, C. E. (1962). Studies on the generality of affective meaning systems. *American Psychologist*, *17*, 10-28.
- Parfit, D. (1984). *Reasons and persons*. Oxford, UK: Clarendon.
- Pelham, B. W., Mirenberg, M. C., & Jones, J. K. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology*, *82*, 469-487.
- Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 83-105). Hillsdale, NJ: Erlbaum.
- Peterson, C., & Seligman, M. E. P. (Eds.). (2004). *Character strengths and virtues*. Washington DC: American Psychological Association.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). New York: Academic Press.
- Piaget, J. (1965). *The moral judgement of the child* (M. Gabain, Trans.). New York: Free Press.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.
- Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychological Review*, *110*, 193-196.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality & Social Psychology*, *67*(4), 741-763.

- Richerson, P. J., & Boyd, R. (1998). The evolution of human ultra-sociality. In I. Eibl-Eibesfeldt & F. K. Salter (Eds.), *Indoctrinability, ideology, and warfare: Evolutionary perspectives* (pp. 71-95). New York: Berghahn.
- Ridley, M. (2004). *Nature via nurture*.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation*. New York: McGraw-Hill.
- Rozin, P. (1982). Human food selection: The interaction of biology, culture and individual experience. In L. M. Barker (Ed.), *The psychobiology of human food selection* (pp. 225-254). Westport, CT: AVI Publishing Co.
- Rozin, P., Fallon, A., & Augustoni-Ziskind, M. L. (1986). The child's conception of food: The development of categories of acceptable and rejected substances. *Journal of Nutrition Education*, 18, 75-81.
- Rozin, P., Fallon, A. E., & Augustoni-Ziskind, M. (1985). The child's conception of food: The development of contamination sensitivity to "disgusting" substances. *Developmental Psychology*, 21, 1075-1079.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The cad triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76, 574-586.
- Saltzstein, H. D., & Kasachkoff, T. (2004). Haidt's moral intuitionist theory. *Review of General Psychology*, 8, 273-282.
- Schnall, S., Haidt, J., & Clore, G. L. (2005). Irrelevant disgust makes moral judgments more severe, for those who listen to their bodies. Unpublished manuscript, university of virginia.
- Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, 58, 878-891.
- Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy*, 2, 307-320.
- Selman, R. (1971). The relation of role taking to the development of moral judgment in children. *Child Development*, 42, 79-91.
- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, 27, No. 187.
- Shweder, R. A., & Haidt, J. (1993). The future of moral psychology: Truth, intuition, and the pluralist way. *Psychological Science*, 4, 360-365.
- Shweder, R. A., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83). Chicago: University of Chicago Press.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119-169). New York: Routledge.
- Singer, P. (1994). *Rethinking life and death*. New York: St. Martin's Press.
- Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Alfred A. Knopf.
- Skitka, L. J. (2002). Do the means always justify the ends, or do the ends sometimes justify the means? A value protection model of justice reasoning. *Personality & Social Psychology Bulletin*, 28, 588-597.
- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39-67). Cambridge, UK: Cambridge University Press.
- Tappan, M. (1997). Language, culture, and moral development: A vygotkian perspective. *Developmental Review*, 17, 78-100.
- Tooby, J., Cosmides, L., & Barrett, C. (in press). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In P. Carruthers (Ed.), *Innateness?*

- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.
- Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, function, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155-243). Chicago: University of Chicago Press.
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9, 160-164.
- Wegner, D., & Bargh, J. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of social psychology (4th ed.)* (pp. 446-496). New York: McGraw Hill.
- Wheatley, T., & Haidt, J. (2005). *The wisdom of repugnance: Hypnotically induced disgust makes moral judgments more severe. (unpublished manuscript, university of virginia.)*
- Wiggins, D. (1987). *Needs, values, truth*. Oxford: Blackwell.
- Wilson, J. Q. (1993). *The moral sense*. New York: Free Press.
- Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Belknap Press.
- Wilson, T. D., Lindsey, S., & Schooler, T. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.
- Wright, R. (1994). *The moral animal*: Pantheon.
- Zahn-Waxler, C., & Radke-Yarrow, M. (1982). The development of altruism: Alternative research strategies. In N. Eisenberg (Ed.), *The development of pro-social behavior*. New York: Academic Press.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.

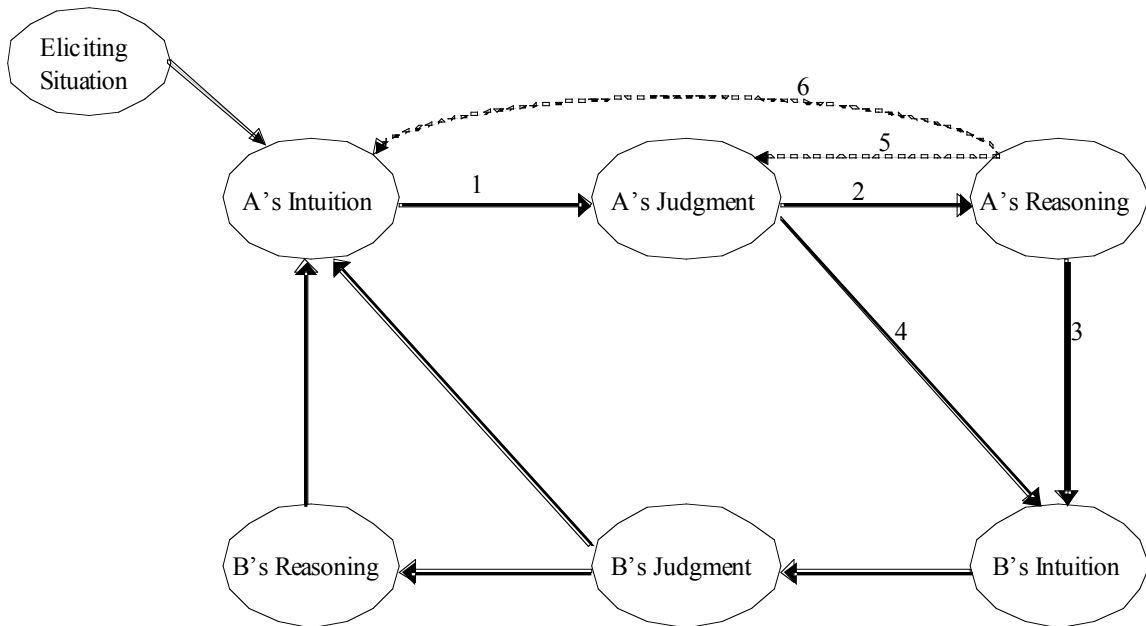


Figure 1. The social intuitionist model of moral judgment. The numbered links, drawn for Person A only, are 1) the intuitive judgment link, 2) the post-hoc reasoning link, 3) the reasoned persuasion link, and 4) the social persuasion link. Two additional links are hypothesized to occur less frequently: 5) the reasoned judgment link, and 6) the private reflection link. (Reprinted from Haidt, 2001)