

**The Emotional Dog and its Rational Tail:  
A Social Intuitionist Approach to Moral Judgment**

Jonathan Haidt  
University of Virginia

October 31, 2000

This is the manuscript that was published, with only minor copy-editing alterations, as:

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*. 108, 814-834

Copyright 2001, American Psychological Association  
To obtain a reprint of the final type-set article, please go through your library's journal services, or contact the author directly

**Abstract**

Research on moral judgment has been dominated by rationalist models, in which moral judgment is thought to be caused by moral reasoning. Four reasons are given for considering the hypothesis that moral reasoning does not cause moral judgment; rather, moral reasoning is usually a post-hoc construction, generated after a judgment has been reached. The social intuitionist model is presented as an alternative to rationalist models. The model is a social model in that it de-emphasizes the private reasoning done by individuals, emphasizing instead the importance of social and cultural influences. The model is an intuitionist model in that it states that moral judgment is generally the result of quick, automatic evaluations (intuitions). The model is more consistent than rationalist models with recent findings in social, cultural, evolutionary, and biological psychology, as well as anthropology and primatology.

**Author notes**

I thank Sara Algoe, Jon Baron, Fredrik Bjorklund, Talbot Brewer, Daniel Fessler, Alan Fiske, Nico Frijda, Chip Heath, Jeff Greenberg, Dacher Keltner, Angeline Lillard, George Loewenstein, Charles Mathewes, Ed Royzman, Paul Rozin, John Sabini, Jonathan Schooler, Stacey Sinclair, Barbara Spellman, Stephen Stich, Stephen Stose, Daniel Wegner, Daniel Willingham, and Timothy Wilson for helpful comments on earlier drafts. Preparation of this paper was supported by the National Institute on Drug Abuse, grant #1-RO3-DA12606-01.

Correspondence concerning this article should be addressed to Jonathan Haidt, Dept. of Psychology, 102 Gilmer Hall Box 400400, University of Virginia, Charlottesville VA 22904-4400. Electronic mail may be sent to haidt at virginia.edu

*Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love?*

Most people who hear the above story immediately say that it was wrong for the siblings to make love, and they then set about searching for reasons (Haidt, Bjorklund, & Murphy, 2000). They point out the dangers of inbreeding, only to remember that Julie and Mark used two forms of birth control. They argue that Julie and Mark will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them. Eventually, many people say something like “I don’t know, I can’t explain it, I just know it’s wrong.” But what model of moral judgment allows a person to know that something was wrong, without knowing why?

Moral psychology has long been dominated by rationalist models of moral judgment (Figure 1). Rationalist approaches in philosophy stress “the power of a priori reason to grasp substantial truths about the world” (Williams, 1967, p.69). Rationalist approaches in moral psychology, by extension, say that moral knowledge and moral judgment are reached primarily by a process of reasoning and reflection (Kohlberg, 1969; Piaget, 1932/1965; Turiel, 1983). Moral emotions such as sympathy may sometimes be inputs to the reasoning process, but moral emotions are not the direct causes of moral judgments. In rationalist models one briefly becomes a judge, weighing issues of harm, rights, justice, and fairness, before passing judgment on Julie and Mark. If no condemning evidence is found, no condemnation is issued.

This paper reviews evidence against rationalist models and proposes an alternative: the social intuitionist model (Figure 2). Intuitionism in

philosophy refers to the view that there are moral truths, and that when people grasp these truths they do so not by a process of ratiocination and reflection, but rather by a process more akin to perception, in which one “just sees without argument that they are and must be true” (Harrison, 1967, p. 72). Thomas Jefferson’s declaration that certain truths are “self-evident” is an example of ethical intuitionism. Intuitionist approaches in moral psychology, by extension, say that moral intuitions (including moral emotions) come first and directly cause moral judgments (Haidt, in press; Kagan, 1984; Shweder & Haidt, 1993; J. Wilson, 1993). Moral intuition is a kind of cognition, but it is not a kind of reasoning.

The social part of the social intuitionist model proposes that moral judgment should be studied as an interpersonal process. Moral reasoning is usually an ex-post facto process used to influence the intuitions (and hence judgments) of other people. In the social intuitionist model one feels a quick flash of revulsion at the thought of incest and one knows intuitively that something is wrong. Then, when faced with a social demand for a verbal justification, one becomes a lawyer trying to build a case, rather than a judge searching for the truth. One puts forth argument after argument, never wavering in the conviction that Julie and Mark were wrong, even after one’s last argument has been shot down. In the social intuitionist model it becomes plausible to say “I don’t know, I can’t explain it, I just know it’s wrong.”

The paper begins with a brief review of the history of rationalism in philosophy and psychology. It then describes the social intuitionist model and recent relevant findings from a variety of fields. These findings offer four reasons for doubting the causality of reasoning in moral judgment: 1) there are two cognitive processes at work -- reasoning and intuition -- and the reasoning process has been overemphasized; 2) reasoning is often motivated; 3) the reasoning process constructs post-hoc justifications, yet we experience the illusion of objective reasoning; and 4) moral action covaries with moral emotion more than with moral reasoning. Since much of this evidence is drawn from research outside of the domain of moral judgment, the social intuitionist model is presented only as a plausible alternative approach to moral psychology, not as an

established fact. The paper therefore concludes with suggestions for future research, and for ways of integrating the findings and insights of rationalism and intuitionism.

It must be stressed at the outset that the social intuitionist model is an anti-rationalist model only in one limited sense: it says that moral reasoning is rarely the direct cause of moral judgment. That is a descriptive claim, about how moral judgments are actually made. It is not a normative or prescriptive claim, about how moral judgments ought to be made. Baron (1998) has demonstrated that people following their moral intuitions often bring about non-optimal or even disastrous consequences in matters of public policy, public health, and the tort system. A correct understanding of the intuitive basis of moral judgment may therefore be useful in helping decision makers avoid mistakes, and in helping educators design programs (and environments) to improve the quality of moral judgment and behavior.

### **Philosophy and the Worship of Reason**

Philosophers have frequently written about the conflict between reason and emotion as a conflict between divinity and animality. Plato's *Timaeus* presents a charming myth in which the gods first created human heads, with their divine cargo of reason, and then found themselves forced to create seething, passionate bodies to help the heads move around in the world. The drama of human moral life was the struggle of the heads to control the bodies by channeling their passions towards virtuous ends. The stoic philosophers took an even dimmer view of the emotions, seeing them as conceptual errors that bound one to the material world and therefore to a life of misery (Solomon, 1993). Medieval Christian philosophers similarly denigrated the emotions because of their link to desire, and hence sin. The 17<sup>th</sup> century's continental rationalists (e.g., Leibniz, Descartes) worshiped reason as much as had Plato, hoping to model all of philosophy on the deductive method developed by Euclid.

But in the 18th century, English and Scottish philosophers (e.g., Shaftesbury, Hutcheson, Hume, and Smith) began discussing alternatives to rationalism. They argued that people have a built-in moral sense which creates pleasurable feelings of

approval toward benevolent acts and corresponding feelings of disapproval towards evil and vice. David Hume in particular proposed that moral judgments are similar in form to aesthetic judgments: they are derived from sentiment, not reason, and we attain moral knowledge by an "immediate feeling and finer internal sense," not by a "chain of argument and induction" (Hume, 1776/1965, p.24). His most radical statement of this position was that "we speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them!" (Hume, 1739/1969, p. 462).

The thrust of Hume's attack on rationalism was that reason alone cannot accomplish the magnificent role it has been given since Plato. Hume saw reason as a tool used by the mind to obtain and process information about events in the world, or about relations among objects. Reason can let us infer that a particular action will lead to the death of many innocent people, but unless we care about those people, unless we have some sentiment that values human life, reason alone cannot advise against taking the action. Hume argued that a person in full possession of reason yet lacking moral sentiment would have difficulty choosing any ends or goals to pursue, and would look like what we now call a psychopath (Hume, 1776/1965; Cleckley, 1955).

Hume's emotivist approach to ethics was not well received by philosophers. Kant's (1785/1959) rationalist ethical theory<sup>2</sup> was created as an attempt to refute Hume, and Kant has had a much larger impact than Hume on modern moral philosophers (e.g., Hare, 1981; Rawls, 1971), many of whom have followed Kant in attempting to deduce a foundation for ethics from the meaning of rationality itself.

### **Psychology and the Focus on Reasoning**

Psychologists, however, freed themselves from the worship of reason in the late 19<sup>th</sup> century, when they abandoned the armchair and went into the laboratory. Until the cognitive revolution of the 1960's, the major schools of psychology did not see reason as the master of anything, and their views on morality were compatible with Hume's emphasis on emotions. Freud (1900/1976) saw people's

judgments as driven by unconscious motives and feelings, which are then rationalized with publicly acceptable reasons. The behaviorists also saw moral reasoning as epiphenomenal in the production of moral behavior, explaining morality as the acts that a society happens to reward or punish (Skinner, 1971).

### Kohlberg and the Cognitive Revolution

But then came Lawrence Kohlberg. Kohlberg's work was a sustained attack on "irrational emotive theories" (1969, p. 188), and his cognitive-developmental theory was an important part of the cognitive revolution. Kohlberg built on Piaget's (1932/1965) pioneering work, developing an interviewing method that was suitable for use with adults as well as children. Kohlberg presented participants with dilemmas in which moral and non-moral claims were present on both sides, and he then looked to see how people resolved the conflict. In his best known dilemma a man named Heinz must decide whether he should break into a druggist's shop to steal a drug that may save the life of his dying wife. Kohlberg found a six level progression of increasing sophistication in how people handled such dilemmas. He claimed that children start as egoists, judging actions by the good or bad consequences they bring to the self. But as children's cognitive abilities expand they develop the ability to "role-take," or see a situation from other people's perspectives. The experience of role-taking drives the child on to the less egocentric and more "adequate" conventional and then post-conventional levels of moral reasoning.

Kohlberg's focus is on development, but in each of his major papers he addresses the question of mechanism. He consistently endorses a rationalist and somewhat Platonic model in which affect may be taken into account by reason (as in Figure 1), but it is reasoning that ultimately makes the decisions:

We are claiming... that the moral force in personality is cognitive. Affective forces are involved in moral decisions, but affect is neither moral nor immoral. When the affective arousal is channeled into moral directions, it is moral; when it is not so channeled, it is not. The moral channeling mechanisms themselves are cognitive (1971, p.230-231).

Kohlberg is quite explicit that the cognitive mechanisms he discusses involve conscious, language based thinking. He is interested in the phenomenology of moral reasoning, and he describes one of the pillars of his approach as the assumption that "moral reasoning is the conscious process of using ordinary moral language" (Kohlberg, Levine, & Hewer, 1983 p. 69).

### After Kohlberg

Kohlberg trained or inspired most of the leading researchers in moral psychology today (see chapters in Kurtines & Gewirtz, 1991; and Lapsley, 1996). Rationalism still rules, and there appears to be a consensus that morality lives within the individual mind as a trait-like cognitive attainment, a set of knowledge structures about moral standards that children create for themselves in the course of their everyday reasoning (see Darley, 1993).

The social-interactionist perspective (Nucci & Turiel, 1978; Turiel, 1983, 1998; Turiel, Killen, & Helwig, 1987) is one of the most widely used approaches at present, and can serve as an illustrative model. This research is based on a method developed by Nucci and Turiel (1978) in which children are interviewed about rule violations. After giving an initial judgment, the child is asked to respond to a series of probe questions designed to assess how the child thinks about the rule in question (e.g., if there was no rule, would the action be OK? Could the rule be changed?). Participants are also asked to provide justifications of their judgments.

In the social-interactionist model, people are said to think about the consequences of an action before determining whether the action is a moral violation. Actions that lead to injustice, to harm, or to the violation of rights are recognized as falling within the moral domain, and are treated differently from other kinds of rule violations. Rules prohibiting moral violations are judged, even by young children, to be universally applicable and unalterable. Actions that involve no injustice, harm, or rights violations are treated as violations of social conventions (involving locally agreed-upon uniformities of behavior within social systems) or as personal issues (areas of individual prerogative).

Researchers in this tradition are sensitive to the ways that moral development occurs in a social context, driven forward by children's interactions

with peers in such contexts as turn-taking, sharing, harming, and responding to harm. This emphasis on social interaction is in harmony with the social part of the social intuitionist model, and is not a source of contention in the present paper. The central source of contention, and the focus of the present paper, concerns the causal role of reflective, conscious reasoning.

### Questioning the Causality of Reasoning

It is undeniable that people engage in moral reasoning. But does the evidence really show that such reasoning is the cause of moral judgment, rather than the consequence? Turiel, Hildebrandt, and Wainryb (1991) examined young adults' reasoning about issues of abortion, homosexuality, pornography, and incest. They found that people who judged the actions to be moral violations also talked about harmful consequences, whereas people who thought the actions were not wrong generally cited no harmful consequences. Turiel et al. interpret these findings as showing the importance of "informational assumptions," for example: people who thought that life begins at conception were generally opposed to abortion, whereas people who thought that life begins later were generally not opposed to abortion. But in making this interpretation Turiel et al. made a jump from correlation to causation. The correlation between judgment and supporting belief does not show that the belief caused the judgment. An intuitionist interpretation is just as plausible: the anti-abortion judgment (a gut feeling that abortion is bad) caused the belief that life begins at conception (an ex post facto rationalization of the gut feeling).

Haidt, Koller, and Dias (1993) found evidence for such an intuitionist interpretation. They examined American and Brazilian responses to actions that were offensive yet harmless, such as eating one's dead pet dog, cleaning one's toilet with the national flag, or eating a chicken carcass one has just used for masturbation. The stories were carefully constructed so that no plausible harm could be found, and most participants directly stated that nobody was hurt by the actions in question. Yet participants still usually said the actions were wrong, and universally wrong. They frequently made statements such as "it's just wrong to have sex with a chicken." Furthermore, their affective reactions to

the stories (statements that it would bother them to witness the action) were better predictors of their moral judgment than were their claims about harmful consequences. Haidt and Hersh (in press) found the same thing when they interviewed conservatives and liberals about sexual morality issues, including homosexuality, incest, and unusual forms of masturbation. For both groups, affective reactions were good predictors of judgment, while perceptions of harmfulness were not. Haidt and Hersh also found that participants were often "morally dumbfounded" (Haidt, Bjorklund, & Murphy, 2000), that is, they would stutter, laugh, and express surprise at their inability to find supporting reasons, yet they would not change their initial judgments of condemnation.

It seems, then, that for affectively charged events such as incest and other taboo violations, an intuitionist model may be more plausible than a rationalist model. But can an intuitionist model handle the entire range of moral judgment? Can it accommodate the findings from rationalist research programs while also explaining new phenomena, and leading to new and testable predictions? The social intuitionist model may be able to do so.

### **The Social Intuitionist Model**

The central claim of the social intuitionist model is that moral judgment is caused by quick moral intuitions, and is followed (when needed) by slow, ex-post facto moral reasoning. Clear definitions of moral judgment, moral intuition, and moral reasoning are therefore needed.

### Moral Judgment

Moral philosophers have long struggled to distinguish moral judgments from other kinds of judgments (e.g., aesthetics, skill, or personal taste). Rather than seeking a formal definition that lists the necessary and sufficient features of a moral judgment, the present paper takes a more empirical approach, starting from a behavioral fact about human beings: that in every society, people talk about and evaluate the actions of other people, and these evaluations have consequences for future interactions (Boehm, 1999). Many of these evaluations occur against the backdrop of specific cultural practices, in which one praises or criticizes

the skills or talents of an individual (e.g., “she is a daring chef”). But an important subset of these evaluations are made with respect to virtues or goods that are applied to everyone in the society (e.g., fairness, honesty, or piety in some cultures), or to everyone in a certain social category (e.g., chastity for young women in some cultures, or generosity for lineage heads). These virtues are obligatory in that everyone (within the relevant categories) is expected to strive to attain them. People who fail to embody these virtues, or whose actions betray a lack of respect for them, are subject to criticism, ostracism, or some other punishment. It is this subset of evaluations that is at issue in the present paper. (For more on moral goods see Ross, 1930; Shweder & Haidt, 1993.)

Moral judgments are therefore defined as: evaluations (good versus bad) of the actions or character of a person that are made with respect to a set of virtues held by a culture or subculture to be obligatory. The definition is intentionally left broad to allow a large gray area of marginally moral judgments. For example, “eating a low fat diet” may not qualify as a moral virtue for most philosophers, yet in health-conscious subcultures, people who eat cheeseburgers and milkshakes are seen as morally inferior to those who eat salad and chicken (Stein & Nemeroff, 1995).

### Moral Reasoning

Everyday moral reasoners are sometimes said to be like scientists, who learn by forming and testing hypotheses, who build working models of the social world as they interact with it, and who consult these models when making moral judgments (Turiel, 1983). A key feature of the scientist metaphor is that judgment is a kind of inference, made in several steps. The reasoner searches for relevant evidence, weighs evidence, coordinates evidence with theories, and reaches 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 now be defined as: conscious mental activity that consists of transforming given information about people in order to reach a moral judgment. 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).

### Moral Intuition

Commentators on intuition have generally stressed the fact that a judgment, solution, or other conclusion appears suddenly and effortlessly in consciousness, without any awareness of the mental processes that led to the outcome (Bastick, 1982; Simon, 1992). Bruner (1960 p.57) said that intuition does not advance in careful steps; rather, it involves “manoeuvres based seemingly on an implicit perception of the total problem. The thinker arrives at an answer, which may be right or wrong, with little if any awareness of the process by which he reached it.” It must be stressed that the contrast of intuition and reasoning is not the contrast of emotion and cognition. Intuition, reasoning, and the appraisals contained in emotions (Frijda, 1986; Lazarus, 1991) are all forms of cognition. Rather, the words “intuition” and “reasoning” are intended to capture the contrast made by dozens of philosophers and psychologists between two kinds of cognition. The most important distinctions (see Table 1) are that intuition occurs quickly, effortlessly, and automatically, such that the outcome but not the process is accessible to consciousness, while reasoning occurs more slowly, requires some effort, and involves at least some steps that are accessible to consciousness.

Building on Bastick, Bruner, Simon, and others, moral intuition can be 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. Moral intuition is therefore the psychological process that the Scottish philosophers talked about, a process akin to aesthetic judgment: one sees or hears about a social

event and one instantly feels approval or disapproval.

### The Links in the Model

The social intuitionist model is composed of four principal links or processes, shown as solid arrows in Figure 2. The existence of each link is well established by prior research in some domains of judgment, although not necessarily in the domain of moral judgment. The model is therefore presented as a proposal to spur thinking and new research on moral judgment.

1) The intuitive judgment link. The model proposes that moral judgments appear in consciousness automatically and effortlessly as the result of moral intuitions. Examples of this link in non-moral cognition include Zajonc's (1980) demonstrations that affectively valenced evaluations are made ubiquitously and rapidly, before any conscious processing has taken place. More recent examples include findings that much of social cognition operates automatically and implicitly (Bargh & Chartrand, 1999; Greenwald & Banaji, 1995).

2) The post-hoc reasoning link. The model proposes that moral reasoning is an effortful process, engaged in after a moral judgment is made, in which a person searches for arguments that will support an already-made judgment. Nisbett and Wilson (1977) demonstrated such post-hoc reasoning for causal explanations. 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-stated hypothesis.

3) The reasoned persuasion link. The model proposes that moral reasoning is produced and sent forth verbally in order to justify one's already-made moral judgment to others. Such reasoning can sometimes affect other people, although moral discussions and arguments are notorious for the rarity with which persuasion takes place. Since moral positions always have an affective component to them, it is hypothesized that reasoned persuasion works not by providing logically compelling arguments, but by triggering new affectively valenced intuitions in the listener. The importance of using affective persuasion to change affectively

based attitudes has been demonstrated by Edwards and von Hippel (1995), and by Shavitt (1990).

4) The social persuasion link. Because people are highly attuned to the emergence of group norms, the model proposes that the mere fact that friends, allies, and acquaintances have made a moral judgment exerts a direct influence on others, even if no reasoned persuasion is used. Such social forces may elicit only outward conformity (Asch, 1956), but in many cases people's privately held judgments are directly shaped by the judgments of others (Berger & Luckman, 1967; Newcomb, 1943; Sherif, 1935).

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. It is hypothesized that people rarely override their initial intuitive judgments just by reasoning privately to themselves, because reasoning is rarely used to question one's own attitudes or beliefs (see the motivated reasoning problem, below).

However, people are capable of engaging in private moral reasoning, and 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. While some of these cases may be illusions (see below, the post-hoc reasoning problem), other cases may be real, particularly among philosophers, one of the few groups that has been found to reason well (Kuhn, 1991). The full social intuitionist model therefore includes two ways in which private reasoning can shape moral judgments.

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.

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.

Rationalist models focus on links 5 and 6. The social intuitionist model, in contrast, says that moral judgment consists primarily of links 1 through 4, although it allows that links 5 and 6 may sometimes contribute (such as during a formal moral judgment interview). The next section of this paper reviews four problems for rationalist models. For each problem, a social intuitionist re-interpretation of the evidence is offered, relying primarily on links 1 through 4.

#### **Four Reasons to Doubt the Causal Importance of Reason**

##### 1) The Dual Process Problem: There Is a Ubiquitous and Understudied Intuitive Process at Work

It is now widely accepted in social and cognitive psychology that there are often two processing systems at work when a person makes judgments or solves problems (see Table 1, and the edited volume by Chaiken & Trope, 1999). Because these two systems typically run in parallel and are capable of reaching differing conclusions, these models are usually called “dual process” models. Dual process models have thus far had little impact on moral judgment research, as most researchers have focused their efforts on understanding the reasoning process (but see Gibbs, 1991; and Eisenberg, Shea, Carlo, & Knight, 1991). Yet there is evidence that moral judgment works like other kinds of judgment, in which most of the action is in the intuitive process.

Automatic evaluation. Affective evaluation occurs so quickly, automatically, and pervasively that it is generally thought to be an integral part of perception. Zajonc (1980) synthesized findings from a variety of fields to create a modern version of Wundt’s (1907) affective primacy theory, in which he argued that feeling and thought are to some extent separate systems with separate biological bases. The affective system has primacy in every sense: it came first in phylogeny, it emerges first in ontogeny, it is triggered more quickly in real-time judgments, and it is more powerful and irrevocable when the two systems yield conflicting judgments (see also Reber, 1993). Research on the automatic evaluation effect confirms that very brief or even subliminal presentations of affectively valenced words (Bargh, Chaiken, Raymond, & Hymes, 1996; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), facial expressions (Murphy & Zajonc, 1993) and photographs of people and animals (Hermans, de Houwer, & Eelen, 1994) alter the time it takes to evaluate a target object presented immediately afterwards, indicating that affective processing is at work within a quarter second of stimulus presentation.

Automatic moral judgment. Moral judgments typically involve more complex social stimuli than the simple words and visual objects used in automatic evaluation studies. Could moral judgments be made automatically as well? The emerging view in social cognition is that most of our behaviors and judgments are in fact made automatically (i.e., without intention, effort, or awareness of process; Bargh, 1994; Bargh & Chartrand, 1999; Greenwald & Banaji, 1995).

The literature most relevant to moral judgment is the literature on attitudes, where a central question has been: how do people form attitudes about other people? The evidence indicates that attitude formation is better described as a set of automatic processes than as a process of deliberation and reflection about the traits of a person. People form first impressions at first sight (Albright, Kenny, & Malloy, 1988), and the impressions that they form from observing a “thin slice” of behavior (as little as 5 seconds) are almost identical to the impressions they form from much longer and more leisurely observation and deliberation (Ambady & Rosenthal, 1992). These first impressions alter subsequent



evaluations, creating a halo effect (Thorndike, 1920), in which positive evaluations of non-moral traits such as attractiveness lead to beliefs that a person possesses corresponding moral traits such as kindness and good character (Dion, Berscheid, & Walster, 1972). People also categorize other people instantly and automatically, applying stereotypes that often include morally evaluated traits (e.g., aggressiveness for African Americans, Devine, 1989). All of these findings illustrate the operation of the intuitive judgment link (link 1 in Figure 2), in which the perception of a person or an event leads instantly and automatically to a moral judgment without any conscious reflection or reasoning.

Another illustration of automatic moral judgment can be seen in the literature on persuasion. Moral discourse in its natural setting is often a kind of persuasion, in which one person tells others about some event, and tries to recruit them to her reading of the event. According to Chaiken's (1987) heuristic-systematic model of persuasion, people are guided in part by the "principle of least effort." Since people have limited cognitive resources, and since heuristic processing is easy and adequate for most tasks, heuristic processing (the intuitive process) is generally used unless there is a special need to engage in systematic processing (see also Simon, 1967). A particularly important heuristic for the study of moral judgment is the "I agree with people I like" heuristic (Chaiken, 1980). If your friend is telling you how Robert mistreated her, there is little need for you to think systematically about the good reasons Robert might have had. The mere fact that your friend has made a judgment affects your own intuitions directly, illustrating the social persuasion link (link 4). It is only if the agreement heuristic leads to other conflicts (e.g., if Robert is a friend of yours) that your "sufficiency threshold" will be raised above your actual level of confidence, triggering effortful systematic processing (links 5 and 6) to close the gap.

However the social intuitionist model posits that moral reasoning is usually done interpersonally rather than privately. If Robert is in fact a friend of yours, then you and your friend might present arguments to each other (link 3, the reasoned persuasion link) in the hope of triggering new intuitions, getting the other to "see" Robert's actions in a better or worse light. Moral discussions can

then be modeled as a repeated cycle through links 1, 2, and 3 in person A, then in person B, then in person A, etc. Link 4 would exert a constant pressure towards agreement if the two parties are friends, and a constant pressure against agreement if the two parties dislike each other. If at least one of the parties begins without a strong initial intuition then some degree of convergence is likely. However if both parties begin with strongly felt opposing intuitions (as in a debate over abortion), then reasoned persuasion is likely to have little effect, except that the post-hoc reasoning triggered in the other person could lead to even greater disagreement, a process labeled "attitude polarization" by Lord, Ross, and Lepper (1979).

Petty and Cacioppo's (1986) elaboration-likelihood model gives a similar reading of the standard moral judgment discussion. If you feel a strong identification with the source of the persuasive message (your friend) and you have no conflict motivating elaborated thinking, then a peripheral process is sufficient to lead to an attitude shift, a judgment that Robert is evil. However if the person talking to you is a stranger (a research psychologist) who challenges your judgment at every turn ("what if Heinz didn't love his wife, should he still steal the drug?"), then you will be forced to engage in extensive effortful, verbal, central processing. Standard moral judgment interviews may therefore create an unnaturally reasoned form of moral judgment, leading to the erroneous conclusion that moral judgment is primarily a reasoning process. And since forcing people to introspect to find reasons for their attitudes can change those attitudes temporarily (Wilson & J. Schooler, 1991; Wilson, Lindsey, & T. Schooler, 2000), standard moral judgment interviews might not even provide a valid measure of people's real moral beliefs. (See also Schooler, Fiore, & Brandimonte, 1997, on the impairments caused by forcing people to verbalize what they know intuitively.)

The social intuitionist solution. The social intuitionist model is fully compatible with modern dual process theories. Like those theories, the model posits that the intuitive process is the default process, handling everyday moral judgments in a rapid, easy, and holistic way. It is primarily when intuitions conflict, or when the social situation

demands thorough examination of all facets of a scenario that the reasoning process is called upon. Reasoning can occur privately (links 5 and 6), and such solitary moral reasoning may be common among philosophers and among those high on need for cognition (Cacioppo & Petty, 1982). Yet ever since Plato wrote his *Dialogues*, philosophers have recognized that moral reasoning naturally occurs in a social setting, between people who can challenge each other's arguments and trigger new intuitions (links 3 and 4). The social intuitionist model avoids the traditional focus on conscious private reasoning, and draws attention to the role of moral intuitions, and of other people, in shaping moral judgments.

## 2) The Motivated Reasoning Problem: The Reasoning Process Is More like a Lawyer Defending a Client than a Judge or Scientist Seeking Truth

It appears, then, that a dual process model may be appropriate for a theory of moral judgment. If so, then the relationship between the two processes must be specified. Is the reasoning process the "smarter" but more cognitively expensive process, called in whenever the intuitive process is unable to solve a problem cheaply? Or is the relationship one of master and servant, as Hume suggested, in which reason's main job is to formulate arguments that support one's intuitive conclusions? Research on motivated reasoning and on everyday reasoning both suggest that the post-hoc reasoning link (link 2) is more important than the reasoned judgment and private reflection links (5 and 6).

There are two major classes of motives that have been shown to bias and direct reasoning. The first class can be called "relatedness motives," for it includes concerns about impression management and smooth interaction with other people. The second class can be called "coherence motives," for it includes a variety of defensive mechanisms triggered by cognitive dissonance and threats to the validity of one's cultural worldview.

Relatedness motives. From an evolutionary perspective it would be strange if our moral judgment machinery was designed principally for accuracy, with no concern for the disastrous effects of periodically siding with our enemies and against our friends. Studies of attitudes, person perception, and persuasion show that desires for harmony and agreement do indeed have strong biasing effects on

judgments. Chaiken and her colleagues incorporated "impression motivation" into the heuristic-systematic model, which is described as "the desire to hold attitudes and beliefs that will satisfy current social goals" (Chen & Chaiken, 1999, p. 78). Chen, Shechter, and Chaiken (1996) found that people who expected to discuss an issue with a partner whose views were known expressed initial attitudes, before the interaction, that were shifted towards those of their anticipated partner. More broadly, Darley and Berscheid (1967) found that people rate a description of a person's personality as more likeable if they expect to interact with the person than if they do not expect to interact.

The existence of motivations to agree with our friends and allies means that we can be directly affected by their judgments (the social persuasion link). The mere fact that your friend expresses a moral judgment against X is often sufficient to cause in you a critical attitude towards X. Such direct influence, circumventing reasoning entirely, fits with Chartrand and Bargh's (1999) recent demonstration of the "chameleon effect," in which people unconsciously mimic the postures, mannerisms, and facial expressions of their interaction partners. Chartrand and Bargh found that such automatic mimicry is socially adaptive, for people who are "in sync" with another person are liked better by that person.

Coherence motives. Psychologists since Freud have argued that people construct views of themselves and of the world, and that they experience potentially crippling anxiety when these constructions are threatened (Moskowitz, Skurnik, & Galinsky, 1999). Research on cognitive dissonance (Festinger, 1957; Wicklund & Brehm, 1976) showed just how readily people change their thinking and beliefs to avoid the threat of internal contradictions. More recently, Chaiken, Giner-Sorolla, and Chen (1996, p.557) defined defense motivation as "the desire to hold attitudes and beliefs that are congruent with existing self-definitional attitudes and beliefs." Self-definitional attitudes include values and moral commitments. When defense motivation is triggered, both heuristic and systematic thinking work to preserve self-definitional attitudes.

The biasing effects of defense motivation can be seen in studies that challenge participants' moral and

political ideology. Lord, Ross, and Lepper (1979) found that students with strong opinions about the death penalty, when exposed to research evidence on both sides of the issue, accepted evidence supporting their prior belief uncritically, while subjecting opposing evidence to much greater scrutiny. Lerner's (1965) "just world" hypothesis stated that people have a need to believe that they live in a world where people generally get what they deserve. People who suffer for no reason are a threat to this belief, so participants adjusted their moral judgments, derogating or blaming innocent victims (Lerner & Miller, 1978). Tetlock et al. (2000) found that people's willingness to use relevant base rate information, or to engage in counterfactual thinking, depended on whether or not their "sacred values" were threatened by doing so. In all of these examples, reasoning is used to defend prior moral commitments.

Moral judgments are also affected by the defensive motivations of terror management (Solomon, Greenberg, & Pyszczynski, 1991). When people are asked to think about their own deaths, they appear to suppress a generalized fear of mortality by clinging more tightly to their cultural world view. Death-primed participants then shift their moral judgments to defend that world view. They mete out harsher punishment to violators of cultural values, and they give bigger rewards to people who behaved morally (Rosenblatt et al., 1989). Death-primed participants have more negative attitudes towards those who do not fully share their world view (e.g., Jews; Greenberg et al., 1990). From a terror management perspective, moral judgment is a special kind of judgment, since moral judgments always implicate the cultural world view. It is plausible to say "I don't like asparagus, but I don't care if you eat it." It is not plausible to say "I think human life is sacred, but I don't care if you kill him."

Mechanisms of bias. Studies of everyday reasoning reveal the mechanisms by which relatedness and coherence motivations make people act like lawyers. Kuhn (1991) found that most people have difficulty understanding what evidence is, and when pressed to give evidence in support of their theories they generally give anecdotes or illustrative examples instead. Furthermore, people show a strong tendency to search for anecdotes and

other "evidence" exclusively on their preferred side of an issue, a pattern that has been called the "my-side bias" (Baron, 1995; Perkins, Farady, & Bushey, 1991). Once people find supporting evidence, even a single piece of bad evidence, they often stop the search, since they have a "makes-sense epistemology" (Perkins, Allen, & Hafner, 1983) in which the goal of thinking is not to reach the most accurate conclusion; it is to find the first conclusion that hangs together well and that fits with one's important prior beliefs.

Research in social cognition also indicates that people often behave like "intuitive lawyers" rather than "intuitive scientists" (Baumeister & Newman, 1994). Kunda's (1990) review of "motivated reasoning" concludes that "directional goals" (motivations to reach a pre-ordained conclusion) work primarily by causing a biased search in memory for supporting evidence only. However Pyszczynski and Greenberg (1987) proposed a more comprehensive "biased hypothesis testing" model, in which self-serving motives bias each stage of the hypothesis testing sequence, including the selection of initial hypotheses, the generation of inferences, the search for evidence, the evaluation of evidence, and the amount of evidence needed before one is willing to make an inference. Research on the "confirmatory bias" (Snyder & Swan, 1978) shows that people do not always seek to confirm their initial hypothesis; sometimes they ask the right questions to get at the truth (Higgins & Bargh, 1987; Trope & Bassok, 1982). However such demonstrations of truth-seeking always involve hypotheses that the participant has no need to defend (e.g., "the person you are about to meet is an extrovert"). When hypotheses involve one's moral commitments (e.g., "the death penalty does not deter murder"), the empirical findings generally show bias and motivated reasoning (Kuhn, 1989; Lord, Ross, & Lepper, 1979).

This review is not intended to imply that people are stupid or irrational. It is intended to demonstrate that the roots of human intelligence, rationality, and ethical sophistication should not be sought in our ability to search for and evaluate evidence in an open and unbiased way. Rather than following the ancient Greeks in worshipping reason, we should instead look for the roots of human intelligence, rationality, and virtue in what the mind does best: perception,

intuition, and other mental operations that are quick, effortless, and generally quite accurate (Gigerenzer & Goldstein, 1996; Margolis, 1987).

The social intuitionist solution. The reasoning process in moral judgment may be capable of working objectively under very limited circumstances: when the person has adequate time and processing capacity, a motivation to be accurate, no a priori judgment to defend or justify, and when no relatedness or coherence motivations are triggered (Forgas, 1995; Wegner & Bargh, 1998). Such circumstances may be found in moral judgment studies using hypothetical and unemotional dilemmas. Rationalist research methods may therefore create an unusual and non-representative kind of moral judgment. But in real judgment situations, such as when people are gossiping or arguing, relatedness motives are always at work. If more shocking or threatening issues are being judged, such as abortion, euthanasia, or consensual incest, then coherence motives will be at work too. Under these more realistic circumstances moral reasoning is not left free to search for truth, but is likely to be hired out like a lawyer by various motives, employed only to seek confirmation of pre-ordained conclusions.

### 3) The Post-hoc Problem: the Reasoning Process Readily Constructs Justifications of Intuitive Judgments, Causing the Illusion of Objective Reasoning

When people are asked to explain the causes of their judgments and actions they frequently cite factors that could not have mattered, and they fail to recognize factors that did matter. Nisbett and Schachter (1966), for example, asked participants to take electric shocks, either with or without a placebo pill that was said to produce the same symptoms as electric shock. Participants in the pill condition apparently attributed their heart palpitations and butterflies in the stomach to the pill, and were able to take four times as much shock as those who had no such misattribution available for their symptoms. Yet when the placebo condition participants were asked if they had made such an attribution, only 25% of them said that they had. The remaining participants denied that they had thought about the pill and instead made up a variety of explanations for their greater shock tolerance, e.g., “Well, I used

to build radios and stuff when I was 13 or 14, and maybe I got used to electric shock” (Nisbett & Wilson, 1977, p.237).

Nisbett and Wilson (1977) interpreted such causal explanations as post-hoc constructions. When asked to explain their behaviors, people engage in an effortful search that may feel like a kind of introspection. But what people are searching for is not a memory of the actual cognitive processes that caused their behaviors, for these processes are not accessible to consciousness. Rather, people are searching for plausible theories about why they might have done what they did. People turn first to a “pool of culturally supplied explanations for behavior,” which Nisbett and Wilson (1977, p.248) refer to as “a priori causal theories.” When asked why he enjoyed a party, a person turns first to his cultural knowledge about why people enjoy parties, chooses a reason, and then searches for evidence that the reason was applicable. The search is likely to be a one-sided search of memory for supporting evidence only (Kunda, 1990; Pyszczynski & Greenberg, 1987).

Additional illustrations of post-hoc causal reasoning can be found in studies in which hypnosis (Zimbardo, LaBerge, & Butler, 1993) and subliminal presentation (Kunst-Wilson & Zajonc, 1980) were used to make people perform actions. When asked to explain their actions or choices, people readily made up reasons that sounded plausible but were false. Split brain patients show this effect in its most dramatic form. When the left hand, guided by the right brain, performs an action, the verbal centers in the left brain readily make up stories to explain it (Gazzaniga, Bogen, & Sperry, 1962). The language centers are so skilled at making up post-hoc causal explanations that Gazzaniga (1985) speaks of an “interpreter” module. He argues that behavior is usually produced by mental modules to which consciousness has no access, but the interpreter module provides a running commentary anyway, constantly generating hypotheses to explain why the self might have performed any particular behavior.

Post-hoc moral reasoning. The idea that people generate causal explanations out of “a priori causal theories” is easily extended into the moral domain. In a moral judgment interview a participant is asked to decide if an action is right or wrong, and is then

asked to explain why she thinks so. But if people have no access to the processes behind their automatic initial evaluations, then how do they go about providing justifications? By consulting their a priori moral theories. A priori moral theories can be defined as a pool of culturally supplied norms for evaluating and criticizing the behavior of others. A priori moral theories provide acceptable reasons for praise and blame (e.g., “unprovoked harm is bad”; “people should strive to live up to God’s commandments”). Since the justifications that people give are closely related to the moral judgments that they make, prior researchers have assumed that the justificatory reasons caused the judgments. But if people lack access to their automatic judgment processes then the reverse causal path becomes more plausible.

If this reverse path is common, then the enormous literature on moral reasoning can be reinterpreted as a kind of ethnography of the a priori moral theories held by various communities<sup>3</sup> and age groups. Kohlberg’s (1969) studies demonstrate that young children in many cultures hold the a priori moral theory that “acts that get punished are wrong; acts that get rewarded are good” (stages 1 and 2), but they soon advance to the theory that “acts that others approve of are good; acts that others condemn are bad” (stage 3). If such statements were the rules that children really used to evaluate actions, then children at stages 1 and 2 would conclude that actions that are not punished must not be bad. Yet Turiel (1983) has shown that young children do not believe this. They say that harmful acts, such as hitting and pulling hair, are wrong whether they are punished or not. They even say that such acts would be wrong if adults ordered them to be done (Damon, 1977; Laupa & Turiel, 1986). So when a child offers the stage 1 statement that “it’s wrong because she’ll get punished,” the child is not introspecting on the reasoning that led to his condemnation; he is just giving a reason that sounds plausible, perhaps a reason he himself has heard from adults (“if you do that, I will punish you”).

The illusions of moral judgment. If moral reasoning is generally a post-hoc construction intended to justify automatic moral intuitions, then our moral life is plagued by two illusions. The first illusion can be called the “wag-the-dog” illusion: we

believe that our own moral judgment (the dog) is driven by our own moral reasoning (the tail). The second illusion can be called the “wag-the-other-dog’s-tail” illusion: in a moral argument, we expect the successful rebuttal of an opponent’s arguments to change the opponent’s mind. Such a belief is like thinking that forcing a dog’s tail to wag by moving it with your hand should make the dog happy.

The wag-the-dog illusion follows directly from the mechanics of the reasoning process described above. Pyszczynski and Greenberg (1987) point out that by going through all the steps of hypothesis testing, even though every step can be biased by self-serving motivations, people can maintain an “illusion of objectivity” about the way they think. The wag-the-dog illusion may therefore be one of the mechanisms underlying naive realism (Griffin & Ross, 1991; Robinson, Keltner, Ward, & Ross, 1995), the finding that people think that they see the world as it is, while their opponents in a moral dispute are biased by ideology and self-interest.

The bitterness, futility, and self-righteousness of most moral arguments can now be explicated. In a debate on abortion, politics, consensual incest, or on what my friend did to your friend, both sides believe that their positions are based on reasoning about the facts and issues involved (the wag-the-dog illusion). Both sides present what they take to be excellent arguments in support of their positions. Both sides expect the other side to be responsive to such reasons (the wag-the-other-dog’s-tail illusion). When the other side fails to be affected by such good reasons, each side concludes that the other side must be closed-minded or insincere. In this way the culture wars over issues such as homosexuality and abortion can generate morally motivated players on both sides who believe that their opponents are not morally motivated (Haidt & Hersh, in press; Robinson, Keltner, Ward, & Ross, 1995).

The social intuitionist solution. People have quick and automatic moral intuitions, and when called upon to justify these intuitions they generate post-hoc justifications out of a priori moral theories. They do not realize that they are doing this, so they fall prey to two illusions. Moral arguments are therefore like shadow boxing matches: each contestant lands heavy blows to the opponent’s shadow, then wonders why he doesn’t fall down. Moral reasoning may therefore have little persuasive

power in conflict situations, but the social intuitionist model says that moral reasoning can be effective in influencing people before a conflict arises. Our words and ideas do affect our friends, allies, and even strangers, via the reasoned-persuasion link. If one can get the other person to “see” the issue in a new way, perhaps by re-framing a problem to trigger new intuitions, then one can influence others with one’s words. Martin Luther King Jr’s “I Have a Dream” speech was remarkably effective in this task, using metaphors and visual images more than propositional logic to get White Americans to see and thus feel that racial segregation was unjust, and un-American (see Lakoff, 1996 on the role of metaphor in political persuasion).

#### 4) the Action Problem: Moral Action Covaries with Moral Emotion More than with Moral Reasoning

The analysis thus far has focused on moral judgment, not moral behavior. But the debate between rationalism and intuitionism can also be carried out using moral action as the dependent variable. There is a literature that directly examines the relationship between moral reasoning and moral action, and there is a literature that examines what happens when moral reasoning and moral emotions become dissociated (in the case of psychopaths).

The weak link between moral reasoning and moral action. In a major review of the literature on moral cognition and action, Blasi (1980) concluded that “moral reasoning and moral action are statistically related” (p.37). But what is the nature of this relationship? Blasi was careful to state that the connection between moral reasoning ability and moral behavior is only a correlation, but later authors in the cognitive developmental tradition read the relationship as causal, stating that higher levels of moral reasoning cause better moral behavior (e.g., Lapsley, 1996). Yet Blasi’s review raised the possibility that a third variable caused both better reasoning and better behavior: intelligence. Blasi found that IQ was consistently related to honesty, and he concluded that future investigators must do a better job of controlling for IQ. Kohlberg (1969) reported that scores on his moral judgment interviews correlated with measures of IQ in the .30 to .50 range. Rest (1979) reports correlations of .20

to .50 between IQ and his Defining Issues Test (DIT).

Intelligence may also be related to better moral behavior by a pathway that does not run through better moral reasoning. Metcalfe and Mischel (1999) proposed a dual process model of willpower in which two separate but interacting systems govern human behavior in the face of temptation. The “hot” system is specialized for quick emotional processing, and it makes heavy use of amygdala-based memory. The “cool” system is specialized for complex spatiotemporal and episodic representation and thought. It relies on hippocampal memory and frontal lobe planning and inhibition areas. It can block the impulses of the hot system, but it develops later in life, making childhood and adolescence seem like a long struggle to overcome impulsiveness and gain self-control. This theory was proposed in part to explain the astonishing finding that the number of seconds pre-schoolers were able to delay choosing an immediate small reward (one marshmallow) in favor of a later, bigger reward (two marshmallows) was a powerful predictor of adolescent social and cognitive competence, including SAT scores and ability to exert self control in frustrating situations, measured about 13 years later (Shoda, Mischel, & Peake, 1990).

The correlation that Blasi found between moral reasoning and moral behavior may therefore be explained by a third variable, the strength of the cool system. Children start off with limited ability to resist temptation, but as the hippocampus and frontal cortex finish their development, children become more able to inhibit impulsive behaviors. Some children start off with a more effective cool system (Kochanska et al., 1996), due to better or faster frontal cortex development. Frontal cortex development makes these children smarter, and they therefore perform better on measures of moral reasoning, but their improved moral behavior comes more from their greater self-regulatory abilities than from their greater moral reasoning abilities. The development of the cool system does not represent the triumph of reasoning over emotion; rather, Metcalfe and Mischel (1999, p.16) see the successful development and integration of the cool system as an essential feature of emotional intelligence.

This re-interpretation is supported by the fact that moral reasoning ability, in Blasi's (1980) review, was most predictive of negative morality -- refraining from delinquent behavior. Criminologists have consistently found an inverse relationship between criminality and IQ. Even after correcting for SES, the difference between delinquent and non-delinquent adolescent populations is approximately eight IQ points (Hirschi & Hindelang, 1977). However the story for positive morality -- directly helping others -- is less clear. Blasi found some support for the claim that high scorers on Kohlberg's and Rest's scales were more likely to help other people, but more recent studies have raised doubts. Hart and Fegley (1995), and Colby and Damon (1992) both compared highly pro-social moral exemplars to non-exemplars, and found that the groups did not differ in their moral reasoning ability assessed by Kohlbergian techniques. A recent review of evidence supporting the utility of the Defining Issues Test described three studies that showed a relationship between DIT scores and negative moral behaviors, but none showing a relationship between DIT scores and positive morality (Thoma, Narvaez, Rest, & Derryberry, 1999).

The relationship between moral reasoning ability and moral behavior therefore appears to be weak and inconsistent, once intelligence is partialled out. Emotional and self-regulatory factors seem to be more powerful determinants of actual behavior (Mischel & Mischel, 1976).

The strong link between moral emotions and moral action. Further evidence that moral reasoning matters less than moral emotions comes from the study of psychopaths. Cleckley's (1955) case studies present chilling portraits of people in whom reasoning has become dissociated from moral emotions. Cleckley characterizes psychopaths as having good intelligence and a lack of delusions or irrational thinking. Psychopaths know the rules of social behavior and they understand the harmful consequences of their actions for others. They simply do not care about those consequences. Cleckley's psychopaths show a general poverty of major affective reactions, particularly those that would be triggered by the suffering of others (remorse, sympathy), condemnation by others (shame, embarrassment), or attachment to others

(love, grief). (See Hare, 1993, for a more recent discussion of the emotional deficit.) Psychopaths can steal from their "friends," dismember live animals, and even murder their parents to collect insurance benefits, without showing any trace of remorse, or of shame when caught. The very existence of the psychopath illustrates Hume's statement that "'tis not contrary to reason to prefer the destruction of the whole world to the scratching of my little finger" (1739/1969, p.461). It is not contrary to reason to kill your parents for money, unless it is also contrary to sentiment.

Several lines of research are converging on the conclusion that psychopaths and people with antisocial personality disorder differ from normal people in the operation of the frontal cortex. Mednick et al. (1982) reviewed studies of EEG differences between criminals and non-criminals, and concluded that the bulk of the research points to differences in the anterior regions of the brain. More recent studies using PET techniques have narrowed the location of interest to the pre-frontal cortex (Raine, 1997). Samples of aggressive offenders show reduced metabolic activity in this area, relative to controls (Raine et al., 1994).

The importance of the prefrontal cortex for moral behavior has been most fully explored by Damasio and his colleagues, who have found a consistent pattern of changes associated with damage to the ventro-medial area of the pre-frontal cortex (VMPFC, the area behind the bridge of the nose). Patients with damage restricted to the VMPFC show no reduction in their reasoning abilities. They retain full knowledge of moral rules and social conventions, and they show normal abilities to solve logic problems, financial problems, and even hypothetical moral dilemmas (Damasio, 1994). Yet when faced with real decisions, they perform disastrously, showing poor judgment, indecisiveness, and what appears to be irrational behavior.

Damasio and his colleagues have demonstrated that the central deficit resulting from destruction of the VMPFC is the loss of emotional responsiveness to the world in general, and to one's behavioral choices in particular. When shown pictures that arouse strong skin conductance responses in normal people (nudity, mutilation, people dying) these patients show no response (Damasio, Tranel, & Damasio, 1990), mirroring the lack of autonomic

responsiveness of psychopaths (Hare & Quinn, 1971). The patients know that the images should affect them, but they report feeling nothing. Damasio refers to this pattern of affect loss combined with intact reasoning as “acquired sociopathy.” Patients with acquired sociopathy do not generally become moral monsters, perhaps since they have a lifetime of normal emotional learning and habit formation behind them, although they do become much less concerned with following social norms, and they sometimes show outrageous and antisocial behavior, as in the case of Phineas Gage (Damasio, 1994). If we imagine a child growing up without a normal VMPFC, who never in his life felt the stings of shame and embarrassment, or the pain of emotional loss or empathic distress, then it becomes almost possible to understand the otherwise incomprehensible behavior of Cleckley’s psychopaths. With no moral sentiments to motivate and constrain them, they simply do not care about the pain they cause and the lives they ruin<sup>4</sup>.

Emotions lead to altruism. If reasoning ability is not sufficient to motivate moral action, then what is? Batson and his colleagues have developed the “empathy-altruism hypothesis,” which states that empathy aroused by the perception of someone’s suffering evokes an altruistic motivation directed toward the ultimate goal of reducing the suffering (Batson, 1987; see also Hoffman, 1982). Batson, O’Quinn, Fuly, Vanderplass, and Isen (1983) found that participants who experienced empathy while watching a woman receiving (fake) electric shocks generally volunteered to take the shocks in her place, even when they were given the option of leaving the scene. Participants who experienced only non-empathic personal distress about the woman’s plight volunteered to trade places with her only when they thought they would have to continue watching the woman receive the shocks. Participants in the first group seemed to be genuinely motivated to help the distressed woman, not to relieve their own distress.

Cialdini and his colleagues have challenged the empathy-altruism hypothesis, using a variety of experimental designs to show that other motives can often explain seemingly altruistic behavior (Cialdini et al., 1987). But throughout this long debate, both sides have consistently agreed that people are often motivated to help others, and that the mechanisms involved in this helping are primarily affective,

including empathy as well as reflexive distress, sadness, guilt and shame (Cialdini, 1991).

The social intuitionist solution. It is easier to study verbal reasoning than it is to study emotions and intuitions, but reasoning may be the tail wagged by the dog. The dog itself may turn out to be moral intuitions and emotions such as empathy and love (for positive morality) and shame, guilt, and remorse, along with emotional self-regulation abilities (for negative morality; see Haidt, in press, for a review and taxonomy of the moral emotions). A dog’s tail is worth studying because dogs use their tails so frequently for communication. Similarly, moral reasoning is worth studying because people use moral reasoning so frequently for communication. But to really understand how human morality works it may be advisable to shift attention away from the study of moral reasoning and towards the study of intuitive and emotional processes.

This concludes the presentation of the four reasons to doubt the importance of reason, and to look more closely at an alternative model, such as the social intuitionist model.

### **The Mechanism of Intuition**

Because intuition is the heart of the social intuitionist model, more must be said about exactly how intuitive moral judgments are made (link 1). Recent work on the importance of bodily experience, as represented in the mind, makes such an account possible.

#### Gut Feelings in the Mind

The somatic marker hypothesis (Damasio, 1994) states that experiences in the world normally trigger emotional experiences that involve bodily changes and feelings. Once the brain is properly tuned up by repeated experiences of such emotional conditioning (e.g., Pavlov, 1927), the brain areas that monitor these bodily changes begin to respond whenever a similar situation arises. It is then no longer necessary for the rest of the body to be involved. At that point, the mere thought of a particular action becomes sufficient to trigger an “as if” response in the brain, in which the person experiences in a weaker form the same bodily feelings that she would experience if she performed the action. The critical job of the VMPFC is to integrate these feelings, or “somatic markers,”



with the person's other knowledge and planning functions, in order to decide quickly upon a response. Damasio's work fits well with research in social psychology on the "affect as information" hypothesis, which demonstrates that people frequently rely on their moods and momentary flashes of feeling as guides when making judgments and decisions (Clare, Schwartz, & Conway, 1994; Loewenstein, Weber, Hsee, & Welch, 1999; Schwarz & Clare, 1983; see also an fMRI finding that such flashes help explain people's varying responses to philosophical moral dilemmas in Greene, Sommerville, Nystrom, Darley, & Cohen, 2000).

Two recent studies have directly manipulated moral judgments by manipulating somatic markers. Batson, Engel, and Fridell (1999) used false physiological feedback to tell participants about their emotional reactions when listening to stories in which the values of either freedom or equality were threatened. When later asked to choose which value should be selected as a theme for a week-long program of events at their university, participants were more likely to choose the value for which they thought they had shown a stronger visceral reaction. Wheatley and Haidt (2000) manipulated somatic markers even more directly. Highly hypnotizable participants were given the suggestion, under hypnosis, that they would feel a pang of disgust when they saw either the word "take" or the word "often". Participants were then asked to read and make moral judgments about six stories that were designed to elicit mild to moderate disgust, each of which contained either the word "take" or the word "often." Participants made higher ratings of both disgust and moral condemnation about the stories containing their hypnotic disgust word. This study was designed to directly manipulate the intuitive judgment link (link 1), and it demonstrates that artificially increasing the strength of a gut feeling increases the strength of the resulting moral judgment.

#### Metaphor and Embodiment

While Damasio focuses on the role of the autonomic nervous system in thinking, Lakoff and Johnson (1999; Lakoff, 1987) have shown how the entire range of physical and emotional experience may underlie our "embodied cognition." By

analyzing how people think and talk about love, politics, morality, and other issues they have shown that nearly all complex thought relies on metaphors, drawn mostly from our experience as physical creatures. For example, because we all have experience with foods that are easily contaminated, we come to equate purity and cleanliness with goodness in the physical domain. We learn from experience that pure substances are quickly contaminated (e.g., by mold, dust, insects) when not guarded, and that once contaminated, it is often difficult to purify them again. These experiences in the physical world then form the basis (in many cultures) of conceptual schemes about moral purity, e.g., that children start off in a state of purity and innocence, but that they can be corrupted by a single exposure to sex, violence, drugs, homosexuality, or the devil (Haidt, Rozin, McCauley, & Imada, 1997; Rozin, Haidt, & McCauley, 2000). Some losses of purity can be rectified with great difficulty (e.g., exorcism after exposure to the devil), while others cannot be rectified at all (e.g., the loss of virginity).

Moral intuition, then, appears to be the automatic output of an underlying, largely unconscious set of interlinked moral concepts. These concepts may have some innate basis (see below), which is then built up largely by metaphorical extensions from physical experience. Metaphors have entailments, and much of moral argument and persuasion involves trying to get the other person to apply the right metaphor. If Saddam Hussein is like Hitler, it follows that he must be stopped. But if Iraq is like Viet Nam, it follows that the United States should not get involved (Spellman & Holyoak, 1992). Such arguments are indeed a form of reasoning, but they are reasons designed to trigger intuitions in the listener.

### **The Origin of Intuitions**

Perhaps because moral norms vary by culture, class, and historical era, psychologists have generally assumed that morality is learned in childhood, and they have set out to discover how morality gets from outside the child to inside. The social intuitionist model takes a different view. It proposes that morality, like language, is a major evolutionary adaptation for an intensely social species, built into multiple regions of the brain and

body, which is better described as emergent than as learned, yet which requires input and shaping from a particular culture. Moral intuitions are therefore both innate and enculturated. The present section describes the ways in which moral intuitions are innate, while the next section describes the ways in which they are shaped by culture during development.

### Primate Proto-Morality

Darwin (1874/1998) believed that the human “moral sense” grew out of the “social instincts” of other animals, and modern primatological research supports him. While all species can be said to follow descriptive rules for behavior with conspecifics, it is primarily the primates that show signs of prescriptive rules, which de Waal (1991) defines as rules that individuals “have learned to respect because of active reinforcement by others” (p. 338). Chimpanzee groups develop and enforce norms for mating, for playing with or touching infants, and for many other forms of interaction. When one individual violates these norms, others will sometimes look to or even get the attention of the individual whose interests have been violated, who may then take action to punish the transgressor (de Waal, 1991). De Waal’s work indicates that prescriptive behavioral norms can emerge and be understood and enforced by chimpanzees, without the benefit of language or language-based reasoning. Language may greatly increase the human use of norms, but the cognitive and emotional machinery of norm creation and norm enforcement was available long before language existed.

It appears, furthermore, that this machinery has been carried forward into the human mind. Alan Fiske (1991, 1992) has identified four underlying models of social cognition that seem to be at work in all human cultures. His first three models fit closely with descriptions of other primates. Fiske’s first model, “communal sharing” involves the linkage of kindness, kinship, and empathic concern for close others that de Waal describes both for chimpanzees (de Waal, 1996) and for bonobos (de Waal & Lanting, 1997). Fiske’s second model, “authority ranking,” describes the ways that power and rank regulate access to resources, but also obligate superiors to protect their subordinates. Such mutual obligations are clear among chimpanzees (de Waal,

1982; Goodall, 1986). Fiske’s third model, “equality matching,” involves the double-edged reciprocal altruism first described by Trivers (1971). Most apes and many monkeys seem remarkably adept at remembering and repaying both favors and slights (de Waal, 1982; 1996). It is only Fiske’s fourth model, “market pricing,” in which ratio values of goods and services must be computed and aggregated across transactions, which seems to be uniquely human (Haslam, 1997). Given so many close parallels between the social lives of humans and chimpanzees, the burden of proof must fall on those who want to argue for discontinuity, i.e., that human morality arose ex nihilo when we developed the ability to speak and reason.

The above considerations are not meant to imply that chimpanzees have morality, nor that humans are just chimps with post-hoc reasoning skills. There is indeed a moral Rubicon that only *Homo Sapiens* appears to have crossed: widespread third party norm enforcement. Chimpanzee norms generally work at the level of private relationships, where the individual that has been harmed is the one that takes punitive action. Yet human societies are marked by a constant and vigorous discussion of norms and norm violators, and by a willingness to expend individual or community resources to inflict punishment, even by those who were not harmed by the violator (Boehm, 1999). Dunbar (1996) has even proposed that language evolved primarily to fulfill the need for gossip. Only with language is it possible to keep track of who did what to whom, who is in, who is out, who can be trusted, and who should be avoided. While the evolution of language and intelligence may have been driven by the Machiavellian benefits they gave to individuals (Byrne & Whiten, 1988), the combination of language and a full theory of mind (Premack & Premack, 1995) made it possible for large groups of non-kin to reap the benefits of cooperation by monitoring each other’s behavior (via gossip), shunning or punishing cheaters, and rewarding team players.

The social intuitionist model fits with this view of the functions of language by including two interpersonal links. Once morality is located in a group’s efforts to solve cooperation and commitment problems (Darwin, 1874/1998; Frank, 1988), it becomes clear that individuals must simultaneously use language to influence others, while being at least

somewhat open to interpersonal influence as specific norms, values, or judgments spread through a community. A group of judges independently seeking truth is unlikely to reach an effective consensus. But a group of people linked together in a large web of mutual influence (an extension of Figure 2 to multiple parties) may eventually settle into a stable configuration, in the same way that a connectionist network reaches a stable equilibrium after several iterations.

### The Externalization of Intuitions

If many moral intuitions (e.g., sympathy, reciprocity, loyalty) are partially built in by evolution then the most important developmental question is not “how do they get in to the child” but rather “how do they get out?” Fiske (1991) argues that social development should be thought of partly as a process of externalization, in which innate cognitive models manifest themselves as a part of normal maturation. He reviews evidence (e.g., Damon, 1975) showing that the four models emerge during development in an invariant sequence: Communal Sharing in infancy, Authority Ranking by age three, Equality Matching around age four, and Market Pricing during middle or late childhood. This is the same sequence in which the models appear to have emerged phylogenetically in the mammalian and primate lineages.

The contrast between internalization and externalization is particularly clear for Equality Matching. Western parents often try to get their young children to share and to play fairly. If moral development was a matter of gradual internalization, or even of reward and punishment, then children’s adherence to principles of fairness would show a gradual increase throughout early childhood. Instead, Fiske argues that children seem relatively insensitive to issues of fairness until around the age of four, at which point concerns about fairness burst forth and are overgeneralized to social situations in which they were never encouraged, and in which they are often inappropriate. This pattern of sudden similarly timed emergence with overgeneralization suggests the maturation of an endogenous ability, rather than the learning of a set of cultural norms. Only after the cognitive model has externalized itself can it be shaped and refined by cultural norms about when and how it should be used.

## **The Development of Intuitions**

Even if moral intuitions are partially innate, children somehow end up with a morality that is unique to their culture or group. There are at least three related processes by which cultures modify, enhance, or suppress the emergence of moral intuitions to create a specific morality: by selective loss, by immersion in custom complexes, and by peer socialization.

### The Selective Loss of Intuitions

The acquisition of phonology provides a useful analogy for the acquisition of morality. Children are born with the ability to distinguish among hundreds of phonemes, but after a few years of exposure to a specific language they lose the ability to make some unexercised phoneme contrasts (Werker & Tees, 1984). Likewise, Ruth Benedict (1934/1959) suggested, we can imagine a great “arc of culture” on which are arrayed all the possible aspects of human functioning. “A culture that capitalized even a considerable proportion of these would be as unintelligible as a language that used all the clicks, all the glottal stops, all the labials....” (p. 24).

Similarly, a culture that emphasized all of the moral intuitions that the human mind is prepared to experience would risk paralysis, as every action triggered multiple conflicting intuitions. Cultures seem instead to specialize in a subset of human moral potential. For example, Shweder’s theory of the “big three” moral ethics (Shweder, Much, Mahapatra, & Park, 1997; see also Jensen, 1997) proposes that moral “goods” (i.e., culturally shared beliefs about what is morally good and valuable) generally cluster into three complexes, or ethics, which cultures embrace to varying degrees: the ethic of autonomy (focusing on goods that protect the autonomous individual, such as rights, freedom of choice, and personal welfare); the ethic of community (focusing on goods that protect families, nations, and other collectivities, such as loyalty, duty, honor, respectfulness, modesty, and self-control); and the ethic of divinity (focusing on goods that protect the spiritual self, such as piety, and physical and mental purity). A child is born prepared to develop moral intuitions in all three ethics, but her local cultural environment generally stresses only one or two of the ethics. Intuitions within culturally

supported ethics become sharper and more chronically accessible (Higgins, 1996), while intuitions within unsupported ethics become weaker and less accessible. Such “maintenance-loss” models have been documented in other areas of human higher cognition. It seems to be a design feature of mammalian brains that much of neural development is “experience expectant” (Black, Jones, Nelson, & Greenough, 1998). That is, there are developmentally timed periods of high neural plasticity, as though the brain “expected” certain types of experience to be present at a certain time to guide its final wiring.

Such sensitive periods are well documented in the development of sensory systems (Hubel & Wiesel, 1970), and language (Johnson & Newport, 1989). Huttenlocher (1994) reports that most synapse selection and elimination in the human cerebral cortex occurs in the first few years, but that in the pre-frontal cortex the period of plasticity is greatly delayed. Synapse selection in the pre-frontal cortex starts later, accelerates in late childhood, and then tails off in adolescence (see also Spear, 2000). Since the pre-frontal cortex is the brain area most frequently implicated in moral judgment and behavior (Damasio, Tranel, & Damasio, 1990; Raine, 1997), this suggests that if there is a sensitive period for moral learning it is likely to be later in childhood than psychoanalysts and most American parents suppose.

But how exactly does a culture choose and emphasize a subset of the available intuitions?

#### Immersion in Custom Complexes

The “custom complex” has recently been proposed as the key construct for understanding development within a cultural context (Shweder, Goodnow, Hatano, LeVine, Markus, & Miller, 1998). The custom complex was originally defined by Whiting and Child (1953, p.27) as consisting of “a customary practice and ... the beliefs, values, sanctions, rules, motives and satisfactions associated with it.” The custom complex captures the idea that cultural knowledge is far more than a set of inherited beliefs about the right and wrong ways of doing things. Cultural knowledge is a complex web of explicit and implicit, sensory and propositional, affective, cognitive, and motoric knowledge (D’Andrade, 1984; Shore, 1996).

Custom complexes are easily found in the moral socialization of children. For example in Orissa, India, many spaces and objects are structured by rules of purity and pollution. Foreigners and dogs may be allowed near the entrance to a temple complex, but only worshipers who have properly bathed may be allowed into the central courtyard (Mahapatra, 1981). In the inner sanctum, where the deity sits, only the Brahmin priest is permitted to enter. Private homes have a similar structure with zones of high purity (the kitchen and the room where the household deity is kept), and lower purity. The human body has a similar structure, in which the head is the zone of highest purity, while the feet are highly polluting. Children in Orissa constantly encounter spaces and bodies structured by purity, and they learn to respect the dividing lines. They learn when to remove their shoes, and how to use their heads and feet in a symbolic language of deference (as when one touches one’s head to the feet of a highly respected person). They develop an intuitive sense that purity and impurity must be kept apart. By participating in these interlinked custom complexes, an Oriya child’s physical embodiment comes to include experiences of purity and pollution. When such children later encounter the intellectual content of the ethics of divinity (e.g., ideas of sacredness, asceticism, and transcendence), their minds and bodies are already prepared to accept these ideas, and their truth feels self-evident (see Lakoff, 1987).

American children, in contrast, are immersed in a different set of practices regarding space and the body, supported by a different ideology. When an American adult later travels in Orissa, he may know how rules of purity and pollution govern the use of space, but he knows these things only in a shallow, factual, consciously accessible way; he does not know these things in the deep cognitive/affective/motoric way that a properly enculturated Oriya knows them.

Fiske (1999) reviews evidence in anthropology that children are taught surprisingly little in most cultures, and that they acquire most of their cultural knowledge and expertise by observing and then imitating the practices of older children and adults. (See also Bandura & Walters, 1963, on imitation and social learning.) Fiske argues that anthropologists have generally underestimated the

importance of motor schemas and implicit knowledge, relying instead on the verbal reports of informants as their primary source of ethnographic data. In other words, there is an asymmetry between how culture gets in to children, and how it gets out to anthropologists. Cultural knowledge gets in largely through non-verbal and non-conscious means, but it gets out through conscious verbal communication. This asymmetry brings the Nisbett and Wilson (1977) problem straight into the heart of anthropology: “Informants pressed to explain practices that they themselves learned by observation, imitation, and participation generally have to make up concepts that have very tenuous, often imaginary relations with the manner in which the informants themselves actually acquired or generate the actions in question” (Fiske, 1999, p. 1; emphasis added).

The importance of practice, repetition, and physical movement for the “tuning up” of cultural intuitions is further demonstrated by Lieberman’s (2000) recent review of the neural substrates of intuition. Lieberman finds that social learning uses some of the same circuits in the basal ganglia that motoric learning does, causing many social skills to become rapid and automatic, like well-learned motor sequences. Social skills and judgmental processes that are learned gradually and implicitly then operate unconsciously, projecting their results into consciousness, where they are experienced as intuitions arising from nowhere. (See also Reber, 1993 on implicit learning, and Clark, 1999, on the underestimated role of the body in cognition.)

The implication of these findings for moral psychology is that moral intuitions are developed and shaped as children behave, imitate, and otherwise take part in the practices and custom complexes of their culture. Participation in custom complexes in this way provides a cultural “front end” for Damasio’s (1994) somatic marker hypothesis, and for Lakoff’s (1987) embodied cognition. Even though people in all cultures have more or less the same bodies, they have different embodiments, and therefore they end up with different minds.

#### Peer Socialization

The social intuitionist model presents people as intensely social creatures whose moral judgments are

strongly shaped by the judgments of those around them. But whose judgments have the strongest effects on children? Harris (1995) pointed out that children’s task in late childhood and adolescence is not to become like their parents but to fit into their peer group, for it is among peers that alliances must be formed and prestige garnered. She therefore proposed a group socialization theory in which children acquire their culture – including moral values – from their peers, just as they acquire their phonology (i.e., children of immigrants copy the accent of their peers, not their parents).

Harris’s emphasis on peers receives support from a study by Minoura (1992) of Japanese children who spent a few years in California when their fathers were transferred to the United States for work. Minoura found that there was a sensitive period for culture learning between the ages of 9 and 15. When children spent a few years in the United States during this period, they developed American ways of interacting with friends, and American ways of feeling about problems in interactions. A few years spent in America before that period led to shallower, non-emotional learning about norms, and left no lasting effects. A few years spent in America after the age of 15 lead to puzzlement and culture-shock, but to little change in the self. These later arrivals, like their parents, knew and could state explicitly the American norms for interpersonal behavior, friendship, and self-promotion, yet these norms did not get internalized. The norms never came to be automatic, or to feel self-evidently valid, as intuitive knowledge would be if acquired during the sensitive period.

Putting together all of the developmental theories and findings presented above yields the following expansion of the social intuitionist model: Moral development is primarily a matter of the maturation and cultural shaping of endogenous intuitions. People can acquire explicit propositional knowledge about right and wrong in adulthood, but it is primarily through participation in custom complexes (Shweder, et al., 1998) involving sensory, motor, and other forms of implicit knowledge (Fiske, 1999; Lieberman, 2000; Shore, 1996), that are shared with one’s peers during the sensitive period of late childhood and adolescence (Harris, 1995; Huttenlocher, 1994; Minoura, 1992) that one comes to feel, physically and emotionally (Damasio, 1994;

Lakoff & Johnson, 1999), the self-evident truth of moral propositions.

### **Integrating Rationalism and Intuitionism**

The debate between rationalism and intuitionism is an old one, but the divide between the two approaches may not be unbridgeable. Both sides agree that people have emotions and intuitions, people engage in reasoning, and people are influenced by each other. The challenge, then, is to specify how these processes fit together. Rationalist models do this by focusing on reasoning and then discussing the other processes in terms of their effects on reasoning. Emotions matter since they can be inputs to reasoning. Social settings and social interactions matter because they encourage or retard the development of reasoning, in part by providing or blocking opportunities for role-taking. But if researchers want to get at the heart of the process, the place where most of the variance is located, they should focus on moral reasoning.

The social intuitionist model proposes a very different arrangement, one that fully integrates reasoning, emotion, intuition, and social influence. The discussion thus far may have given the impression that the model dismisses reasoning as post-hoc rationalization (link 2). However it must be stressed that four of the six links in the model are reasoning links, and three of these links (3, 5, and 6) are hypothesized to have real causal effects on moral judgment.

Link 3, the reasoned persuasion link, says that people's (ex post facto) moral reasoning can have a causal effect – on other people's intuitions. In the social intuitionist view moral judgment is not just a single act that occurs in a single person's mind. It is an ongoing process, often spread out over time and over multiple people. Reasons and arguments can circulate and affect people, even if individuals rarely engage in private moral reasoning for themselves.

Link 6, the reflective judgment link, allows that people may sometimes engage in private moral reasoning for themselves, particularly when their initial intuitions conflict. Abortion may feel wrong to many people when they think about the fetus, but right when their attention shifts to the woman. When competing intuitions are evenly matched the judgment system becomes deadlocked and the

“master” (in Hume's metaphor) falls silent. Under such circumstances one may go through repeated cycles of links 6, 1, and 2, using reasoning and intuition together to break the deadlock. That is, if one consciously examines a dilemma, focusing in turn on each party involved, various intuitions will be triggered (link 6), leading to various contradictory judgments (link 1). Reasoning can then be used to construct a case to support each judgment (link 2). If reasoning is more successful building a case for one of the judgments than the others, it will begin to feel right, and there will be less temptation (and ability) to consider additional points of view. This is an account of how a “makes sense” epistemology (Perkins, Allen, & Hafner, 1983) may become a “feels right” ethics. We use conscious reflection to mull over a problem until one side feels right. Then we stop.

Link 5, the reasoned judgment link, recognizes that a person could, in principle, simply reason her way to a judgment that contradicts her initial intuition. The literature on everyday reasoning (Kuhn, 1991) suggests that such an ability may be common only among philosophers, who have been extensively trained and socialized to follow reasoning even to very disturbing conclusions (as in the case of Socrates, or the more recent works of Derek Parfit and Peter Singer). Yet the fact that there are at least a few people among us who can reach such conclusions on their own, and then argue for them eloquently (link 3), means that pure moral reasoning can play a causal role in the moral life of a society.

If the social intuitionist model is correct as a description of human moral judgment, it may be possible to use the model to get reasoning and intuition working more effectively together in real moral judgments. One approach would be to directly teach moral thinking and reasoning skills, thereby encouraging people to use links 5 and 6 more often. However attempts to directly teach thinking and reasoning in a classroom setting generally show little transfer to activities outside of the classroom (Nickerson, 1994), and since moral judgment involves “hotter” topics than are usually dealt with in such courses, the degree of transfer is likely to be even smaller.

A more intuitionist approach is to treat moral judgment style as an aspect of culture, and to try to

create a culture that fosters a more balanced, reflective, and fair-minded style of judgment. The “Just Community” schools that Kohlberg created in the 1970's (Power, Higgins, & Kohlberg, 1989) appear to do just that. By making high school students create their own rules, enforce their own discipline, and vote on numerous policies, Kohlberg created an environment where students enacted democracy. By putting students and teachers on an equal footing (all had just one vote; all used first name only; all sat in a circle on the floor at community meetings), Kohlberg created an environment where students and teachers enacted equality. Years of such implicit learning, coupled with explicit discussion, should gradually tune up intuitions (Fiske, 1999; Lieberman, 2000) about justice, rights, and fairness, leading perhaps to an automatic tendency to look at problems from multiple perspectives. By creating a community in which moral talk was ubiquitous (link 3, reasoned persuasion), and in which adults modeled good moral thinking, Kohlberg may well have strengthened his students’ tendency to use link 6 (private reflection) on their own. (See Baron, 2000, for more on how cultural beliefs and practices about thinking can help or hinder good thinking.)

The social intuitionist model also offers more general advice for improving moral judgment. If the principal difficulty in objective moral reasoning is the biased search for evidence (Kunda, 1990; Perkins, Farady, & Bushey, 1991), then people should take advantage of the social persuasion link (link 4) and get other people to help them improve their reasoning. By seeking out discourse partners who are respected for their wisdom and open-mindedness, and by talking about the evidence, justifications, and mitigating factors involved in a potential moral violation, people can help trigger a variety of conflicting intuitions in each other. If more conflicting intuitions are triggered, the final judgment is likely to be more nuanced and ultimately more reasonable.

The social intuitionist model, therefore, is not an anti-rationalist model. It is a model about the complex and dynamic ways that intuition, reasoning, and social influences interact to produce moral judgment.

### Testing the Social Intuitionist Model

The social intuitionist model is more complex and comprehensive than most rationalist models. Is the extra complexity necessary? Does the model do a better job of explaining and illuminating human moral life? That is a question that future research must decide. There are at least three kinds of research that may shed light on the relative merits of the model.

1) Interfering with reasoning. If reasoning is a slow and effortful process that demands attentional resources, while intuition is fast, effortless, and undemanding (see Table 1), then manipulations that interfere with reasoning during a moral judgment interview should affect the quality of the post-hoc reasoning produced, without affecting the quality of the initial judgment. Rationalist models, in contrast, predict that the quality and speed of a judgment should be heavily dependent on one’s reasoning ability.

2) Ecological variation. This paper has suggested that standard moral judgment interviews represent unique and ecologically suspect settings in which a variety of factors conspire to maximize the amount and quality of reasoning. If this is true then the reasoning produced in such interviews is consistent both with rationalist models and with the private reflection loop of the social intuitionist model (links 1, 2, and 6). However as the conditions of the interview are gradually changed to increase ecological validity, the social intuitionist model predicts that the reasoning produced should become recognizably post-hoc. Alterations that would increase ecological validity include using real (rather than hypothetical) stories, asking about people known to the participant, working questions into a normal conversation (not a formal interview), and conducting the conversation in front of other people (not alone in a private room). Post-hoc reasoning can be recognized by three features: 1) attempts to change facts about the story or to introduce new and tangential concerns, 2) a lack of responsiveness of the judgment to large changes in the facts of the story, and 3) a longer delay between the time the evaluation is made and the time that the first substantive reason is produced.

3) Consilience. Edward O. Wilson (1998) resurrected the term “consilience” to refer to the degree to which facts and theories link up across disciplines to create a common groundwork of explanation. He argued that theories that contribute to the unification of the sciences should be preferred to those that contribute to their fragmentation. The present paper has tried to show that the social intuitionist model easily links findings in social and developmental psychology to recent findings and theories in neuroscience, primatology, and anthropology. But perhaps a similar case can be made for rationalist models. The debate between rationalism and intuitionism, now over 200 years old, is not just a debate between specific models; it is a debate between perspectives on the human mind. All of the disciplines that study the mind should contribute to the debate.

### **Conclusion**

Rationalist models made sense in the 1960's and 1970's. The cognitive revolution had opened up new ways of thinking about morality and moral development, and it was surely an advance to think about moral judgment as a form of information processing. But times have changed. Now we know (again) that most of cognition occurs automatically and outside of consciousness (Bargh & Chartrand, 1999), and that people cannot tell us how they really reached a judgment (Nisbett & Wilson, 1977). Now we know that the brain is a connectionist system that tunes up slowly, but is then able to evaluate complex situations quickly (Bechtel & Abrahamsen, 1991). Now we know that emotions are not as irrational (Frank, 1988), that reasoning is not as reliable (Kahneman & Tversky, 1984), and that animals are not as amoral (de Waal, 1996) as we thought in the 1970's. The time may be right, therefore, to take another look at Hume's perverse thesis: that moral emotions and intuitions drive moral reasoning, just as a surely as a dog wags its tail.



### Footnotes

1. This is one of Hume's most radical statements, taken from his first book, A Treatise of Human Nature. His more mature work, An Enquiry Concerning the Principles of Morals, raises reason from a slave to a respected assistant of the moral sense, yet it maintains the basic position that "the ultimate ends of human actions can never... be accounted for by reason, but recommend themselves entirely to the sentiments and affections" (p. 131).
2. Kant responded to Hume's skepticism about the powers of reason. He argued that any rational agent could and should figure out the morally correct thing to do by applying the categorical imperative: "I should never act in such a way that I could not also will that my maxim should be a universal law" (1785/1959, p.18).
3. An ironic example of an a priori moral theory used in a post-hoc way is found in Miller's (1999) recent review of the norm of self-interest. Americans strongly embrace the theory that people act, and ought to act, primarily out of self-interest. Americans therefore frequently make up self-interest explanations for their attitudes, votes, and charitable actions, even in cases where they appear to be acting against their self-interest (see also Baron, 1997).
4. In fact, two of the only such children ever studied sound uncannily like Cleckley's psychopaths (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999). See also Grandin's (1995) discussion of how the emotional deficits of autism made it difficult for her to understand many social and moral rules, although her feelings of empathy, particularly for animals, and her feelings of social anxiety, appear to have been a sufficient foundation for a moral compass.

### References

- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. Journal of Personality and Social Psychology, *55*, 387-395.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological Bulletin, *111*, 256-274.
- Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. Nature Neuroscience, *2*, 1032-1037.
- Asch, S. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. Psychological Monographs, *70*, (Whole No. 416.)
- Bandura, A., & Walters, R. (1963). Social learning and personality development. New York: Holt, Rinehart and Winston.
- 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., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditionally automatic activation with a pronunciation task. Journal of Experimental Social Psychology, *32*, 185-210.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. American Psychologist, *54*, 462-479.
- Baron, J. (1995). Myside bias in thinking about abortion. Thinking and Reasoning, *1*, 221-235.
- Baron, J. (1997). The illusion of morality as self-interest. Psychological Science, *8*, 330-335.
- Baron, J. (1998). Judgment misguided: Intuition and error in public decision making. New York: Oxford.
- Baron, J. (2000). Thinking and deciding, 3rd Edition. Cambridge, UK: Cambridge University Press.
- Bastick, T. (1982). Intuition: How we think and act. Chichester, UK: John Wiley.
- Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic? Advances in Experimental Social Psychology, *20*, 65-122.
- Batson, C. D., Engel, C. L., & Fridell, S. R. (1999). Value judgments: Testing the somatic-marker hypothesis using false physiological feedback. Personality and Social Psychology Bulletin, *25*, 1021-1032.
- Batson, C. D., O'Quinn, K., Fulty, J., Vanderplass, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. Journal of Personality and Social Psychology, *45*, 706-718.
- Baumeister, R. F., & Newman, L. S. (1994). Self regulation of cognitive inference and decision processes. Personality and Social Psychology Bulletin, *20*, 3-19.
- Bechtel, W., & Abrahamsen, A. (1991). Connectionism and the mind: An introduction to parallel processing in networks. Cambridge, MA: Blackwell.
- Benedict, R. (1959). Patterns of culture. Boston: Houghton Mifflin.
- Berger, P. L., & Luckman, T. (1967). The social construction of reality. New York: Doubleday.
- Black, J. E., Jones, T. A., Nelson, C. A., & Greenough, W. T. (1998). Neuronal plasticity and the developing brain. In N. E. Alessi, J. T. Coyle, S. I. Harrison, & S. Eth (Eds.), Handbook of child and adolescent psychiatry. (pp. 31-53). New York: John Wiley.
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. Psychological Bulletin, *88*, 1-45.
- Boehm, C. (1999). Hierarchy in the forest: The evolution of egalitarian behavior. Cambridge, MA: Harvard University Press.
- Bruner, J. S. (1960). The process of education. Cambridge, MA: Harvard University Press.
- Bruner, J. S. (1986). Actual minds, possible worlds. Cambridge, MA: Harvard University Press.
- Byrne, R., & Whiten, A. (Eds.). (1988). Machiavellian intelligence. Oxford: Oxford University Press.

- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. Journal of Personality and Social Psychology, 42, 116-131.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 39, 752-766.
- Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), Social influence: The Ontario symposium. (pp. 3-39). Hillsdale, NJ: Erlbaum.
- Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. M. Gollwitzer & J. A. Bargh (Ed.), The psychology of action: Linking cognition and motivation to behavior. (pp. 553-578). New York: Guilford.
- Chaiken, S., & Trope, Y. (Eds.). (1999). Dual process theories in social psychology. New York: Guilford.
- 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.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), Dual process theories in social psychology. (pp. 73-96). New York: Guilford.
- Chen, S., Shechter, D., & Chaiken, S. (1996). Getting at the truth or getting along: Accuracy- versus impression-motivated heuristic and systematic processing. Journal of Personality and Social Psychology, 71, 262-275.
- Cialdini, R. B. (1991). Altruism or egoism? That is (still) the question. Psychological Inquiry, 2, 124-126.
- Cialdini, R., Schaller, M., Houlihan, D., Arps, K., Fultz, J., & Beaman, A. (1987). Empathy based helping: Is it selflessly or selfishly motivated? Journal of Personality and Social Psychology, 52, 749-758.
- Clark, A. (1999). Being there: Putting brain, body, and world together again. Cambridge, MA: MIT Press.
- Cleckley, H. (1955). The mask of sanity. St. Louis: C. V. Mosby.
- Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer & T. K. Srull (Eds.), Handbook of social cognition. (pp. 323-417). Hillsdale, NJ: Erlbaum.
- Colby, A., & Damon, W. (1992). Some do care: Contemporary lives of moral commitment. New York: Free Press.
- Damasio, A. (1994). Descartes' error: Emotion, reason, and the human brain. New York: G. P. Putnam's Sons.
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. Behavioral Brain Research, 41, 81-94.
- Damon, W. (1975). Early conceptions of positive justice as related to the development of logical operations. Child Development, 46, 301-312.
- Damon, W. (1977). The social world of the child. San Francisco: Jossey-Bass.
- D'Andrade, R. G. (1984). Cultural meaning systems. In R. A. Shweder & R. A. LeVine (Eds.), Culture theory. Cambridge: Cambridge University Press.
- Darley, J. (1993). Research on morality: Possible approaches, actual approaches. Psychological Science, 4, 353-357.
- Darley, J. M., & Berscheid, E. (1967). Increased liking as a result of anticipation of personal contact. Human Relations, 20, 29-40.
- Darwin, C. (1874/1998). The descent of man. Amherst, NY: Prometheus Books.
- de Waal, F. (1982). Chimpanzee politics. New York: Harper and Row.
- de Waal, F. (1991). The chimpanzee's sense of social regularity and its relation to the human sense of justice. American Behavioral Scientist, 34, 335-349.
- de Waal, F. (1996). Good natured: The origins of right and wrong in humans and other animals. Cambridge, Mass: Harvard University Press.
- de Waal, F., & Lanting, F. (1997). Bonobo: The forgotten ape. Berkeley, CA: University of California Press.

- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5-18.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. Journal of Personality and Social Psychology, 24, 207-213.
- Dunbar, R. (1996). Grooming, gossip, and the evolution of language. Cambridge, MA: Harvard University Press.
- Edwards, K., & Von Hippel, W. (1995). Hearts and minds: The priority of affective versus cognitive factors in person perception. Personality and Social Psychology Bulletin, 21, 996-1011.
- Eisenberg, N., Shea, C. L., Carlo, G., & Knight, G. P. (1991). Empathy-related responding and cognition: A "chicken and the egg" dilemma. In W. M. Kurtines & J. L. Gewirtz (Eds.), Handbook of moral behavior and development, Vol. 1: Theory. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49, 709-724.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic evaluation of attitudes. Journal of Personality and Social Psychology, 50, 229-238.
- Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford 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.
- Fiske, A. P. (1999). Learning culture the way informants do: Observing, imitating, and participating. (Unpublished manuscript, UCLA Dept. of Anthropology).
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). Psychological Bulletin, 117, 39-66.
- 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.) New York: Norton. (Original work published 1900).
- Frijda, N. (1986). The emotions. Cambridge: Cambridge University Press.
- Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. Psychological Bulletin, 105, 331-351.
- Gazzaniga, M. S. (1985). The social brain. New York: Basic Books.
- Gazzaniga, M. S., Bogen, J. E., & Sperry, R. W. (1962). Some functional effects of sectioning the cerebral commissures in man. Proceedings of the National Academy of Sciences, USA, 48, 1765-1769.
- Gibbs, J. C. (1991). Toward an integration of Kohlberg's and Hoffman's theories of morality. In W. M. Kurtines & J. L. Gewirtz (Eds.), Handbook of moral behavior and development: Advances in theory, research, and application (Vol. 1). (pp. 183-222). Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review, 103, 650-669.
- Goodall, J. (1986). The chimpanzees of Gombe: Patterns of behavior. Cambridge, MA: Belknap Press, Harvard University Press.
- Grandin, T. (1995). Thinking in pictures: and other reports from my life with autism. New York: Doubleday.
- Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A. V., M., Kirkland, S., & Lyon, D. (1990). Evidence for terror management theory II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview. Journal of Personality and Social Psychology, 58, 308-318.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2000). An fMRI study of emotional engagement in moral judgment. Unpublished manuscript, Princeton University.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition. Psychological Review, 102, 4-27.
- Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. Advances in Experimental Social Psychology, 24, 319-359.

- Haidt, J. (in press). The moral emotions. In R. J. Davidson, K. Scherer, & H. H. Goldsmith (Eds.), Handbook of affective sciences. Oxford: Oxford University Press.
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. (Unpublished manuscript, University of Virginia).
- Haidt, J., & Hersh, M. (in press). Sexual morality: the cultures and reasons of liberals and conservatives. Journal of Applied Social Psychology.
- 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., Rozin, P., McCauley, C. R., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. Psychology and Developing Societies, *9*, 107-131.
- Hare, R. D. (1993). Without conscience. New York: Pocket Books.
- Hare, R. D., & Quinn, M. J. (1971). Psychopathy and autonomic conditioning. Journal of Abnormal Psychology, *77*, 223-235.
- Hare, R. M. (1981). Moral thinking: Its levels, method, and point. Oxford: Oxford University Press.
- Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. Psychological Review, *102*, 458-489.
- Harrison, J. (1967). Ethical objectivism. In P. Edwards (Ed.), The Encyclopedia of Philosophy, Vol. 3 & 4. (pp. 71-75). New York: Macmillan.
- Hart, D., & Fegley, S. (1995). Prosocial behavior and caring in adolescence: Relations to self-understanding and social judgment. Child Development, *66*, 1346-1359.
- Haslam, N. (1997). Four grammars for primate social relations. In J. A. Simpson & D. T. Kenrick (Eds.), Evolutionary social psychology. (pp. 297-316). Mahwah, NJ: Lawrence Erlbaum.
- Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic evaluation of evaluative information in memory. Cognition and Emotion, *8*, 515-533.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), Social psychology: Handbook of basic principles. (pp. 133-168). New York: Guilford.
- Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. Annual Review of Psychology, *38*, 369-425.
- Hirschi, T., & Hindelang, M. J. (1977). Intelligence and delinquency: a revisionist view. American Sociological Review, *42*, 571-87.
- Hoffman, M. L. (1982). Development of prosocial motivation: Empathy and guilt. In N. Eisenberg (Ed.), The development of prosocial behavior. New York: Academic Press.
- Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure. Journal of Physiology, *206*, 419-436.
- Hume, D. (1965). An enquiry concerning the principles of morals. Notre Dame: University of Notre Dame Press. (Original work published 1777).
- Hume, D. (1969). A treatise of human nature. London: Penguin. (Original work published 1739 & 1740).
- Huttenlocher, P. R. (1994). Synaptogenesis, synapse elimination, and neural plasticity in human cerebral cortex. In C. A. Nelson (Ed.), Threats to optimal development: Integrating biological, psychological, and social risk factors. The Minnesota symposia on child development, Vol. 27. (pp. 35-54). Hillsdale, NJ: Lawrence Erlbaum.
- Jensen, L. A. (1997). Different worldviews, different morals: America's culture war divide. Human Development, *40*, 325-344.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive Psychology, *21*, 60-99.
- Kagan, J. (1984). The nature of the child. New York: Basic Books.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. American Psychologist, 341-350.

- Kant, I. (1959). Foundation of the metaphysics of morals. (Lewis White Beck, Trans.) Indianapolis: Bobbs-Merrill. (Original work published 1785).
- Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. Child Development, *67*, 490-507.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), Handbook of socialization theory and research. (pp. 347-480). 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.), Cognitive development and epistemology. New York: Academic Press.
- Kohlberg, L., Levine, C., & Hewer, A. (1983). Moral stages: A current formulation and a response to critics. In J. A. Meacham (Ed.), Contributions to human development, Vol. 10. Basel: Karger.
- 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.
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. Science, *207*, 557-558.
- Kurtines, W. M., & Gewirtz, J. L. (Eds.). (1991). Handbook of moral behavior and development (Vols. 1-3). Hillsdale, NJ: Erlbaum.
- Lakoff, G. (1987). Women, fire, and dangerous things. Chicago: University of Chicago Press.
- Lakoff, G. (1996). Moral politics: What conservatives know that liberals don't. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh. New York: Basic Books.
- Lapsley, D. K. (1996). Moral psychology. Boulder, CO: Westview.
- Laupa, M., & Turiel, E. (1986). Children's conceptions of adult and peer authority. Child Development, *57*, 405-412.
- Lazarus, R. S. (1991). Emotion and adaptation. New York: Oxford University Press.
- Lerner, M. J. (1965). Evaluation of performance as a function of performer's reward and attractiveness. Journal of Personality and Social Psychology, *1*, 355-360.
- Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. Psychological Bulletin, *85*, 1030-1051.
- Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. Psychological Bulletin, *126*, 109-137.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, E. S. (1999). Risk as feeling (Unpublished manuscript, Carnegie Mellon University).
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality & Social Psychology, *37*, 2098-2109.
- Mahapatra, M. (1981). Traditional structure and change in an Orissan temple. Calcutta: Punthi Pustak.
- Margolis, H. (1987). Patterns, thinking, and cognition. Chicago: University of Chicago Press.
- Mednick, S. A., Pollock, V., Volavka, J., & Gabrielli, W. F. (1982). Biology and violence. In M. Wolfgang & N. A. Weiner (Eds.), Criminal violence. (pp. 21-80). Beverly Hills: Sage.
- Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. Psychological Review, *106*, 3-19.
- Miller, D. T. (1999). The norm of self-interest. American Psychologist, *54*, 1053-1060.
- Mischel, W., & Mischel, H. N. (1976). A cognitive social-learning approach to morality and self-regulation. In T. Lickona (Ed.), Moral development and behavior: Theory, research, and social issues. (pp. 84-107). New York: Holt, Rinehart and Winston.
- 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.

- Moskowitz, G. B., Skurnik, I., & Galinsky, A. D. (1999). The history of dual process notions, and the future of pre-conscious control. In S. Chaiken & Y. Trope (Eds.), Dual process theories in social psychology. (pp. 12-36). New York: Guilford.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. Journal of Personality and Social Psychology, *64*, 723-729.
- Newcomb, T. M. (1943). Personality and social change: Attitude formation in a student community. New York: Dryden.
- 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., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Schacter, S. (1966). Cognitive manipulation of pain. Journal of Experimental Social Psychology, *2*, 227-236.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, *84*, 231-259.
- Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. Child Development, *49*, 400-407.
- Pavlov, I. (1927). Conditioned reflexes. Oxford: Oxford University Press.
- Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), Thinking: The frontier expands. (pp. 177-189). Hillsdale, NJ: Erlbaum.
- 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.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), Advances in experimental social psychology. (pp. 123-205). New York: Academic Press.
- Piaget, J. (1965). The moral judgement of the child. (M. Gabain, Trans.) New York: Free Press. (Original work published 1932).
- Plato (1949). Timaeus. (Benjamin Jowett, Trans.) Indianapolis: Bobbs-Merrill. (Original work published 4th Cent. BC).
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium. (pp. 55-85). Hillsdale, NJ: Erlbaum.
- Power, F. C., Higgins, A., & Kohlberg, L. (1989). Lawrence Kohlberg's approach to moral education. New York: Columbia University Press.
- Premack, D., & Premack, A. J. (1995). Origins of human social competence. In M. S. Gazzaniga (Ed.), The cognitive neurosciences. (pp. 205-218). Cambridge, MA: MIT Press.
- Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. Advances in Experimental Social Psychology, *20*, 297-340.
- Raine, A. (1997). Antisocial behavior and psychophysiology: A biosocial perspective and a prefrontal dysfunction hypothesis. In D. M. Stoff, J. Breiling, & J. D. Maser (Eds.), Handbook of antisocial behavior. (pp. 289-304). New York: John Wiley.
- Raine, A., Buchsbaum, M. S., Stanley, J., Lottenberg, S., Abel, L., & Stoddard, J. (1994). Selective reductions in pre-frontal glucose metabolism in murderers. Biological Psychiatry, *36*, 365-373.
- Rawls, J. (1971). A theory of justice. Cambridge, MA: Harvard University Press.
- Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious. New York: Oxford University Press.
- Rest, J. R. (1979). Development in judging moral issues. Minneapolis: University of Minnesota Press.

- Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. Journal of Personality and Social Psychology, *68*, 404-417.
- Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. Journal of Personality and Social Psychology, *57*, 681-690.
- Ross, W. D. (1930). The right and the good. Oxford: Oxford University Press.
- Rozin, P., Haidt, J., & McCauley, C. R. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), Handbook of emotions, 2nd Edition. (pp. 637-653). New York: Guilford Press.
- Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. The Psychology of Learning and Motivation, *37*, 291-340.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Information and directive functions of affective states. Journal of Personality and Social Psychology, *45*, 513-523.
- Selman, R. (1971). The relation of role taking to the development of moral judgement in children. Child development, *42*, 79-91.
- Shavitt, S. (1990). The role of attitude objects in attitude formation. Journal of Experimental Social Psychology, *26*, 124-148.
- Sherif, M. (1935). A study of some social factors in perception. Archives of Psychology, *2*.
- Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. Developmental Psychology, *26*, 978-986.
- Shore, B. (1996). Culture in mind: Cognition, culture, and the problem of meaning. New York: Oxford University Press.
- Shweder, R. A., Goodnow, J., Hatano, G., LeVine, R. A., Markus, H., & Miller, P. (1998). The cultural psychology of development: One mind, many mentalities. In W. Damon (Ed.), Handbook of child psychology, 5th Ed., Vol. 1: R. M. Lerner (Ed.), Theoretical models of human development. (pp. 865-937). New York: Wiley.
- 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., 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.
- Simon, H. A. (1967). Motivational and emotional controls of cognition. Psychological Review, *74*, 29-39.
- Simon, H. A. (1992). What is an "explanation" of behavior? Psychological Science, *3*, 150-161.
- Skinner, B. F. (1971). Beyond freedom and dignity. New York: Alfred A. Knopf.
- Snyder, M. & Swann, W. B. (1978). Hypothesis testing processes in social interaction. Journal of Personality and Social Psychology, *36*, 1202-1212.
- Solomon, R. C. (1993). The philosophy of emotions. In M. Lewis & J. Haviland (Eds.), Handbook of emotions. (pp. 3-15). New York: Guilford Press.
- Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). A terror management theory of social behavior. In M. Zanna (Ed.), Advances in experimental social psychology. (pp. 93-159). San Diego: Academic Press.
- Spear, L. P. (2000). Neurobehavioral changes in adolescence. Current Directions in Psychological Science, *9*, 111-114.
- Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. Journal of Personality & Social Psychology, *62*, 913-933.
- Stein, R., & Nemeroff, C. J. (1995). Moral overtones of food: Judgments of others based on what they eat. Personality and Social Psychology Bulletin, *21*, 480-490.
- Tappan, M. (1997). Language, culture, and moral development: A Vygotskian perspective. Developmental Review, *17*, 78-100.



- Tetlock, P. E., Kristel, O. V., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. Journal of Personality and Social Psychology, *78*, 853-870..
- Thoma, S. J., Narvaez, D., Rest, J., & Derryberry, P. (1999). Does moral judgment development reduce to political attitudes or verbal ability? Evidence using the defining issues test. Educational Psychology Review, *11*, 325-341.
- Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of Applied Psychology, *4*, 25-29.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. Quarterly Review of Biology, *46*, 35-57.
- Trope, Y., & Bassok, M. (1982). Information gathering strategies in hypothesis testing. Journal of Experimental Social Psychology, *43*, 22-34.
- Turiel, E. (1983). The development of social knowledge: Morality and convention. Cambridge: Cambridge University Press.
- Turiel, E. (1998). The development of morality. In W. Damon (Ed.), Handbook of child psychology, 5th Ed., Vol. 3: N. Eisenberg (Ed.), Social, emotional, and personality development. (pp. 863-932). New York: Wiley.
- Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Judging social issues: Difficulties, inconsistencies, and consistencies. Monographs of the Society for Research in Child Development, *56*, 1-103.
- 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.
- Wegner, D. (1994). Ironic processes of mental control. Psychological Review, *101*, 34-52.
- 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.
- Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, *7*, 49-63.
- Wheatley, T., & Haidt, J. (2000). Hypnotically induced disgust makes moral judgments more severe. (Manuscript in preparation, University of Virginia.)
- Whiting, J., & Child, I. (1953). Child training and personality. New Haven: Yale University Press.
- Wicklund, R. A., & Brehm, J. W. (1976). Perspectives on cognitive dissonance. Hillsdale, NJ: Erlbaum.
- Williams, B. (1967). Rationalism. In P. Edwards (Ed.), The Encyclopedia of Philosophy, Vol. 7 & 8. (pp. 69-75). New York: Macmillan.
- Wilson, E. O. (1998). Consilience: The unity of knowledge. New York: Alfred A. Knopf.
- Wilson, J. Q. (1993). The moral sense. New York: Free Press.
- Wilson, T. D. (In Press). Strangers to ourselves: Self-insight and the adaptive unconscious. Harvard University Press.
- Wilson, T. D., Lindsey, S., & Schooler, T. (2000). A model of dual attitudes. Psychological Review, *107*, 101-126.
- Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. Journal of Personality and Social Psychology, *60*, 181-192.
- Wundt, W. (1907). Outlines of psychology. Leipzig, Germany: Wilhelm Englemann.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, *35*, 151-175.
- Zimbardo, P. G., LaBerge, S., & Butler, L. D. (1993). Psychophysiological consequences of unexplained arousal: A posthypnotic suggestion paradigm. Journal of Abnormal Psychology, *102*, 466-473.

Table 1  
General Features of the Two Systems

<b>The Intuitive system</b>	<b>The Reasoning System</b>
Fast and effortless	Slow and effortful
Process is unintentional and runs automatically	Process is intentional and controllable
Process is inaccessible; only results enter awareness	Process is consciously accessible and viewable
Does not demand attentional resources	Demands attentional resources, which are limited
Parallel distributed processing	Serial processing
Pattern matching; thought is metaphorical, holistic	Symbol manipulation; thought is truth preserving, analytical
Common to all mammals	Unique to humans over age 2, and perhaps some language-trained apes
Context dependent	Context independent
Platform dependent (depends on the brain and body that houses it)	Platform independent (the process can be transported to any rule following organism or machine)

Note. These contrasts are discussed in Bruner (1986); Chaiken (1980); Epstein (1994); Freud (1900/1976); Margolis (1987); Metcalfe and Mischel (1999); Petty & Cacioppo (1986); Posner and Snyder (1975); Pyszczynski and Greenberg (1987); Reber (1993); Wegner (1994); Wilson (in press); and Zajonc (1980).

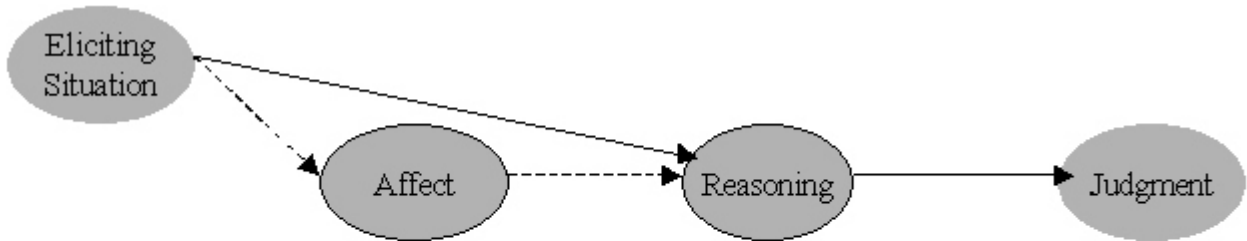


Figure 1.

The rationalist model of moral judgment. Moral affects such as sympathy may sometimes be inputs to moral reasoning

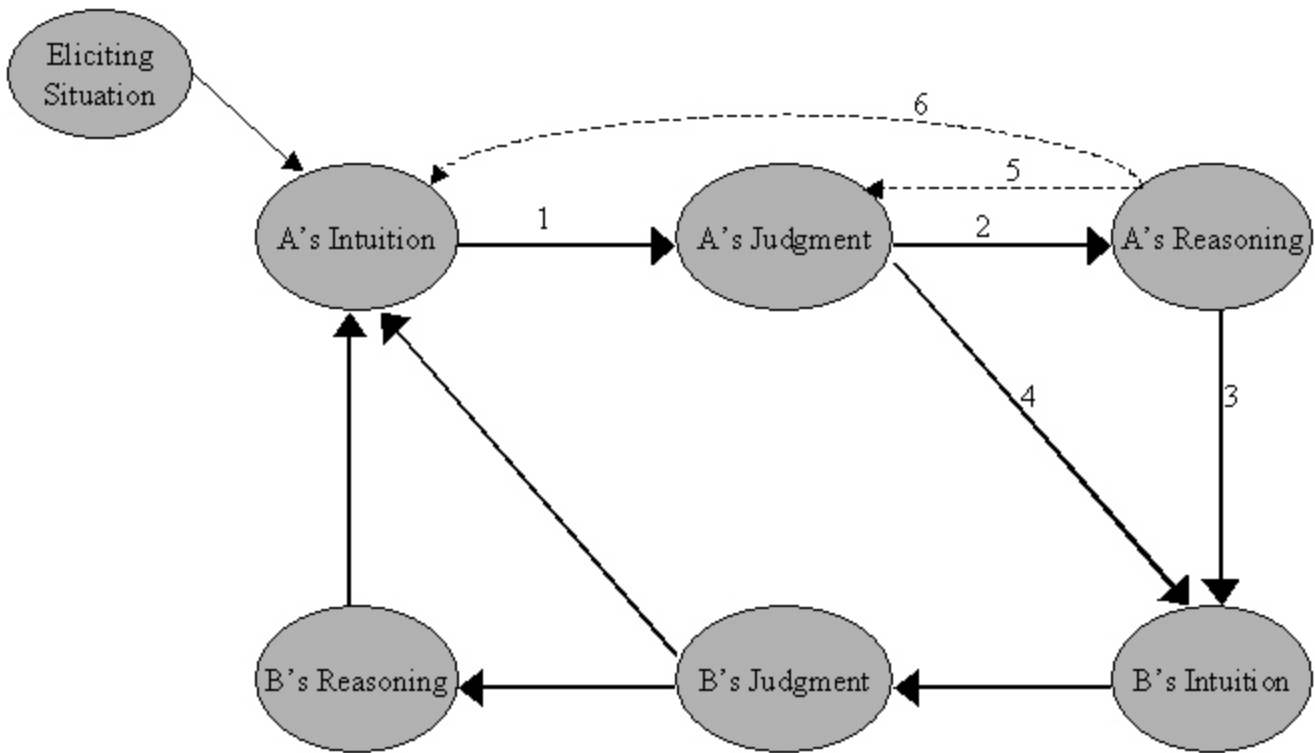


Figure 2. The social intuitionist model of moral judgment. The numbered links, drawn for Person A only, are 1) the intuitive judgment link, 2) the post-hoc reasoning link, 3) the reasoned persuasion link, and 4) the social persuasion link. Two additional links are hypothesized to occur less frequently: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