Moral Psychology

Volume 2: The Cognitive Science of Morality: Intuition and Diversity

edited by Walter Sinnott-Armstrong

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

Contents

Introduction

Walter Sinnott-Armstrong

Moral judgments, emotions, and actions can be studied in many ways. One method cites patterns in observations as evidence of evolutionary origins. That method was exemplified in the chapters of the first volume in this collection. A second method uses similar patterns in observations as evidence of cognitive processes employed in forming the moral judgments, emotions, or actions. That method will be exemplified in the papers in this volume. It can be described as cognitive science.

Cognitive scientists need not say anything about how their proposed cognitive processes evolved in order to reach fascinating and important results. Nonetheless, no cognitive model can be acceptable if it is incompatible with what we know about evolution. Moreover, a cognitive model can derive additional support from a plausible story about how a cognitive process evolved to where it is today. In these ways and more, the papers on cognitive science in this volume are connected to the views on the evolution of morality that were canvassed in the preceding volume.

Similarly, although cognitive scientists do not have to mention brain mechanisms, connections to known brain mechanisms can validate a postulated cognitive model, and any cognitive model must be rejected if it conflicts with what we know about the brain and how it works. The cognitive science of moral judgment, emotion, and action cannot be done thoroughly in isolation from brain science. Hence, the papers on cognitive science in this volume are also connected to those on neuroscience in the following volume.

Nonetheless, this particular volume focuses on cognitive science rather than on evolution or on brain science. The various chapters here illustrate the diversity of approaches within cognitive science.

One very common view in cognitive science claims that the mind often works by means of heuristics—fast and frugal procedures for forming beliefs, reaching decisions, and performing actions. A major debate about

heuristics concerns their reliability. Daniel Kahneman, Amos Tversky, and their collaborators emphasize that heuristics often lead to biases and at least apparent irrationality. In contrast, Gerd Gigerenzer is well-known for emphasizing the usefulness and reliability of heuristics in proper circumstances. In his chapter in this volume, Gigerenzer applies his general approach to moral decisions and intuitions and argues that moral intuitions are nothing more than fast and frugal heuristics of a special sort. He claims that this view is much more realistic than competing views, including some forms of consequentialism, whose standards of rationality cannot be met by real cognitive agents.

Cass Sunstein comments that even good heuristics can produce a kind of irrationality, and he argues that prescriptive treatments of moral heuristics should avoid controversial assumptions about morality. Julia Driver and Don Loeb then claim that Gigerenzer has misinterpreted traditional consequentialism. In his reply, Gigerenzer admits that heuristics produce some errors, but so do all available policies, so heuristics can still reduce errors more than any realistic alternative in some environments. According to Gigerenzer, what we need to study is when—that is, in which environments—specific heuristics lead to better results than any realistic alternative.

Since heuristics predict and explain framing effects, Gigerenzer's chapter leads right into the next chapter, which is by Walter Sinnott-Armstrong and is on framing effects in moral judgment. Framing effects are, basically, variations in beliefs as a result of variations in wording and order. Sinnott-Armstrong emphasizes that, when such variations in wording and order cannot affect the truth of beliefs, framing effects signal unreliability. Hence, if framing effects are widespread enough in moral intuitions, moral believers have reason to suspect that those intuitions are unreliable. Sinnott-Armstrong cites empirical evidence that framing effects are surprisingly common in moral judgments, so he concludes that moral believers have reason to suspect that their moral intuitions are unreliable. This result creates a need for confirmation and thereby undermines traditional moral intuitionism as a response to the skeptical regress problem in moral epistemology.

William Tolhurst and Russ Shafer-Landau both defend moral intuitionism against Sinnott-Armstrong's argument. Tolhurst argues that Sinnott-Armstrong's grounds support only a suspension of belief about whether moral intuitions are reliable or unreliable. Shafer-Landau canvasses different ways of understanding Sinnott-Armstrong's argument and concludes that none of them undermine traditional moral intuitionism. In reply,

Sinnott-Armstrong reformulates his argument in ways that are intended to avoid the criticisms of Tolhurst and Shafer-Landau.

Moral intuitions are also often understood by means of analogy to Noam Chomsky's universal grammar, an analogy suggested by John Rawls.[1] This analogy is developed in chapter 3, by Marc Hauser, Liane Young, and Fiery Cushman. This research group uses a very large Web-based survey as well as data on brain-damaged patients[2] to adjudicate among four models of moral judgment.[3] They conclude that at least some forms of moral judgment are universal and mediated by unconscious and inaccessible principles. This conclusion supports the analogy to linguistics and suggests that these principles could not have been learned from explicit teaching.

In his comment on Hauser et al., Ron Mallon argues that no evidence points to a specialized moral faculty or supports a strong version of the linguistic analogy. Jesse Prinz then emphasizes several disanalogies between morality and language and shows how to account for the data of Hauser et al. without admitting that morality is innate. In their reply, Hauser, Young, and Cushman specify how Mallon and Prinz have misread their linguistic analogy; then they discuss some new evidence for their model.

One of the psychological models criticized by Hauser, Young, and Cushman was presented by Jonathan Haidt and claims that initial moral intuitions are emotional, whereas moral reasoning normally comes later and serves social purposes. Since its original formulation, Haidt's social intuitionist model of moral judgment has been subjected to numerous criticisms.[4] In their wide-ranging chapter here, Haidt and Fredrik Bjorklund develop their view in new ways, summarize the evidence for their view, and respond to criticisms. They discuss where moral beliefs and motivations come from, how moral judgment works, how morality develops, and why moral beliefs vary. In their philosophical conclusion, they argue against monistic theories and for an anthropocentric view of moral truth that is supposed to be neither relativistic nor skeptical insofar as it shows how some moral codes can be superior to others.[5]

Both commentators suggest that Haidt and Bjorklund's theory is more relativistic than Haidt and Bjorklund admit. Daniel Jacobson also argues that social intuitionism in moral psychology coheres best with a kind of sentimentalism in metaethics, rather than with Haidt and Bjorklund's anthropocentric view. Darcia Narvaez then claims that social intuitionism addresses only a small sample of moral judgment, reasoning, and decision. Haidt and Bjorklund reply that their topic is moral judgment rather than moral decision making and which judgments get classified as moral

judgments is under dispute. They also point out how their model can account for more moral reasoning than their critics realize and can avoid extreme versions of moral relativism.

Emotions are also central to morality according to Shaun Nichols, who summarizes his recent book, *Sentimental Rules*, in the next chapter.[6] Nichols holds that emotions play a key role in everyday moral judgment, but emotions cannot be the whole story, since normal people have similar emotional reactions to harms that are not caused by immoral actions. Nichols also argues that philosophical attempts to analyze moral judgments in terms of emotions, such as Allan Gibbard's prominent version of sentimentalism, conflict with recent studies of moral judgments by young children. Nichols concludes that core moral judgment depends on a body of rules—a normative theory. This postulated theory is used to explain recent observations of young children, as well as the relation of moral judgments to conventional judgments and rules regarding disgusting actions.

In his comment, Justin D'Arms defends his own version of sentimentalism against Nichols's arguments and counters that sentimentalists can explain moral disagreement better than Nichols can. James Blair then grants the force of Nichols's arguments against Blair's previous views but questions whether Nichols's postulated normative theory is needed for all moral judgments and decisions, and he suggests that the normative theory's work might be done by the theory of mind, since an agent's intention marks the difference between what is harmful and what is morally wrong. In response, Nichols argues that intention is not enough to explain the difference between being harmful and being wrong and that the only kind of disagreement that D'Arms can explain that he can't is fundamental moral disagreement among nonobjectivists, which is dubious for independent reasons.

Several preceding essays already mentioned moral disagreement, but this crucial topic is explored directly in the next chapter, by John Doris and Alexandra Plakias.[7] Doris and Plakias argue that intractable moral disagreements create serious problems for moral realism. Then they canvass a variety of evidence for such fundamental moral disagreements, including studies of attitudes toward honor and violence in the American South and a new study of Chinese responses to punishing innocent people. Doris and Plakias list possible defusing explanations (ignorance, partiality, irrationality, and prior theoretical commitments) and argue that none of these explanations defuses the disagreement in the cases under discussion, so the moral disagreement is truly fundamental and, hence, creates trouble for moral realism.

Brian Leiter questions whether moral disagreements can be explained away by partiality or irrationality, and he suggests that the history of philosophy (especially Friedrich Nietzsche) is enough by itself to reveal fundamental moral disagreements. Paul Bloomfield then argues that fundamental moral disagreements do not undermine moral realism, because moral realists can accommodate and explain moral divergence, and because widespread moral agreement can exist alongside the documented moral disagreements. In their reply, Plakias and Doris compare disagreements among philosophers with disagreements among everyday people, and disagreements over principles with disagreements over particular cases, as problems for moral realism.

The following chapter, by Don Loeb, also concerns moral realism, but he approaches moral realism from the perspective of the semantics of moral language. Loeb argues that moral language should be studied empirically and that, when we see how moral language actually works, it looks like moral vocabulary contains too much semantic incoherence for moral terms to refer to real properties, as moral realists claim. The crucial incoherence arises from the observation that ordinary people use moral language both to make factual assertions and also to do something incompatible with making such assertions.

Michael Gill suggests that these different uses of moral language are confined to isolated areas of moral thought, so variability need not yield incoherence or undermine moral realism about some areas of morality. Geoffrey Sayre-McCord then asks whether semantic theories in metaethics might be trying to capture not what ordinary people think and say but only what "we" think and say, where the pronoun "we" identifies a group that can properly be seen as making genuine moral judgments. Loeb replies by questioning whether Gill can divide moral language and thought into semantically insulated pockets and whether Sayre-McCord can nonarbitrarily identify a group that makes genuine moral judgments and avoids incoherence.

Morality includes not only rules about which acts are morally wrong but also standards of when agents are responsible for doing such acts. One common claim is that an agent is morally responsible for a harm only if that agent's action or omission caused that harm. Julia Driver defends this standard in the final chapter of this volume. Driver responds to both philosophical and empirical challenges to this standard and suggests that our causal judgments are based on prior judgments of whether the purported cause is unusual either statistically or normatively. Driver also argues that, while psychological research into the folk concepts can be

interesting and helpful, it cannot replace more traditional methods of philosophy.

In response, Joshua Knobe, who was criticized by Driver, and Ben Fraser report the results of two new experiments that are supposed to undermine Driver's suggestion that causal judgments follow judgments of what is unusual or atypical, along with the alternative suggestion that Knobe's previous empirical results can be explained by conversational pragmatics. John Deigh then argues that Driver's claim that moral responsibility entails causation is refuted by legal cases, including one where medical personnel at a jail release a psychotic killer and another where every member of a group is complicit in and, hence, responsible for a harm that only some members of that group caused. Driver replies that her theory is compatible with Knobe and Fraser's new findings and also with Deigh's examples, if criminal responsibility is distinguished from moral responsibility.

These brief summaries cannot, of course, do justice to the rich empirical detail, careful philosophical arguments, and variety of profound issues that arise in these chapters. All together, these chapters show how much contemporary cognitive science has to contribute to moral theory.

Notes

1. This linguistic analogy is discussed further in the chapters by Sripada and Prinz in the first volume of this collection.

2. Hauser's studies of moral judgments by people with brain damage are relevant to the debate between Kennett and Roskies in the third volume of this collection.

3. Two of these models are defended by Haidt and Bjorklund in this volume and by Greene in the third volume of this collection.

4. Compare the chapter by Greene in the third volume of this collection.

5. A similar view is developed and defended by Flanagan, Sarkissian, and Wong in the first volume of this collection.

6. Moral emotions are also discussed by Moll et al. and by Greene in the third volume of this collection.

7. The extent and importance of moral disagreement are also discussed in the chapters by Sripada and Prinz in the first volume of this collection.

1 Moral Intuition = Fast and Frugal Heuristics?

Gerd Gigerenzer

Ordinary Men

On July 13, 1942, the men of Reserve Police Battalion 101, stationed in Poland, were wakened at the crack of dawn and driven to the outskirts of a small Polish village. Armed with additional ammunition, but with no idea what to expect, the 500 men gathered around their well-liked commander, Major Wilhelm Trapp (Browning, 1993). Nervously, Trapp explained that he and his men had been assigned a frightfully unpleasant task, not to his liking, but the orders came from the highest authorities. There were some 1,800 Jews in the village, who were said to be involved with the partisans. The order was to take the male Jews of working age to a work camp. The women, children, and elderly were to be shot on the spot. As he spoke, Trapp had tears in his eyes and visibly fought to control himself. He and his men had never before been confronted with such a task. Concluding his speech, Trapp made an extraordinary offer: If any of the older men did not feel up to the task that lay before them, *they could step out*.

Trapp paused for a moment. The men had a few seconds to decide. A dozen men stepped forward. The others went on to participate in the massacre. Many of them, however, after they had done their duty once, vomited or had other visceral reactions that made it impossible to continue killing and were then assigned to other tasks. Almost all were horrified and disgusted by what they were doing. Yet why did only a mere dozen men out of 500 declare themselves unwilling to participate in the mass murder?

One might first think of anti-Semitism. That, however, is unlikely, as the historian Christopher Browning (1993) documents in his seminal book *Ordinary Men*. Most of the battalion members were middle-aged family men, considered too old to be drafted into the German army, and drafted instead into the police battalion. By virtue of their age, their formative

years had taken place in the pre-Nazi era, and they knew different political standards and moral norms. They came from the city of Hamburg, by reputation one of the least nazified cities in Germany, and from a social class that had been anti-Nazi in its political culture. These men would not have seemed to be a promising group of mass murderers on behalf of the Nazi vision.

The extensive interviews with the men indicate that the primary reason was not conformity with authority either. Unlike in the Milgram experiment, where an authoritative researcher told students to apply electric shocks to other people, Major Trapp explicitly allowed for "disobedience." The men who stepped out experienced no sanctions from him. If neither anti-Semitism nor fear of authority was the explanation, what else had turned ordinary men into mass killers? The documents collected on this case reveal a different reason. Most policemen's behavior seemed to follow a social heuristic:

Don't break ranks.

The men felt "the strong urge not to separate themselves from the group by stepping out" (Browning, 1993, p. 71), even if this conformity meant violating the moral imperative "Don't kill innocent people." For most, it was easier to shoot rather than to break ranks. Browning ends his book with a disturbing question: "Within virtually every social collective, the peer group exerts tremendous pressures on behavior and sets moral norms. If the men of Reserve Police Battalion 101 could become killers under such circumstances, what group of men cannot?" From a moral point of view, nothing can justify this behavior. In trying to understand why certain situations can promote or inhibit morally significant actions, however, we can find an explanation in social heuristics.[1]

Organ Donors

Since 1995, some 50,000 people in the United States have died waiting for a suitable organ donor (Johnson & Goldstein, 2003). Although most Americans say they approve of organ donation, relatively few sign a donor card. Here neither peer pressure, nor obedience, nor fear of being punished seems to be at issue. Why are only 28% of Americans but a striking 99.9% of French citizens donors? Do Americans fear that if emergency room doctors know that the patients are potential organ donors, they won't work as hard to save them? Or are Americans more anxious about a postmortem opening of their bodies than the French? Yet why are only 17% of British citizens but 99.9% of Hungarians donors?

If moral behavior is the result of deliberate moral reasoning, then the problem might be that Americans and the British are not aware of the need for organs. This view calls for an information campaign to raise people's awareness so that they change their behavior. Dozens of such campaigns have been launched in the United States and the United Kingdom with limited success. If moral behavior is the result of stable preferences, as postulated by rational choice theory, then Americans and the British might simply find too little utility in donation. Yet that does not seem to be the case either. Something stronger than preferences and deliberate reasoning appears to guide behavior. The differences between nations seem to be produced by a simple rule, the *default rule*:

If there is a default, do nothing about it.

In explicit-consent countries such as the United States and the United Kingdom, the law is that nobody is a donor without registering to be one. You need to opt in. In presumed-consent countries such as France and Hungary, everyone is a donor unless they opt out. The majority of citizens in these and other countries seem to follow the same default rule, and the striking differences between nations result as a consequence. However, not everyone follows the default rule. Among those who do not, most opt in but few opt out. The 28% of Americans who opted in and the 0.1% of French citizens who opted out illustrate this asymmetry. The perceived rationale behind the rule could be that the existing law is interpreted as a reasonable recommendation; otherwise it would not have been chosen by the policymakers. From a rational choice perspective, however, the default should have little effect because people will override the default if it is not consistent with their preference. After all, one only needs to sign a form to opt in or to opt out. However, the empirical evidence demonstrates that it is the default rule rather than alleged preferences that explains most people's behavior.

Fast and Frugal Heuristics

The two examples illustrate the general thesis of this essay: Morally significant actions (moral actions, for short) can be influenced by simple heuristics. The resulting actions can be morally repulsive, as in the case of mass killing, or positive, as when people donate organs or risk their lives to save that of another person. The underlying heuristic, however, is not good or bad per se.

The study of heuristics will never replace the need for moral deliberation and individual responsibility, but it can help us to understand which

environments influence moral behavior and how to possibly modify them to the better. One and the same heuristic can produce actions we might applaud *and* actions we condemn, depending on where and when a person relies on it. For instance, the don't-break-ranks heuristic can turn a soldier simultaneously into a loyal comrade and into a killer. As an American rifleman recalls about comradeship during World War II: "The reason you storm the beaches is not patriotism or bravery. It's that sense of not wanting to fail your buddies. There's sort of a special sense of kinship" (Terkel, 1997, p. 164). Similarly, the default rule can turn a person into an organ donor or none. What appears as inconsistent behavior—how can such a nice guy act so badly, and how can that nasty person be so nice?—can result from the same underlying heuristic.

In this essay, I will look at moral actions through the lens of the theory of fast and frugal heuristics (Gigerenzer, Todd, & the ABC Research Group, 1999; Gigerenzer & Selten, 2001; Payne, Bettman, & Johnson, 1993). This theory is based on the work on bounded rationality by Nobel laureates Herbert Simon and Reinhard Selten. A heuristic is called "fast" if it can make a decision within little time, and "frugal" if it searches for only little information. The science of heuristics centers on three questions:

1. *Adaptive toolbox* What heuristics do people have at their disposal? What are their building blocks, and which evolved (or learned) abilities do these exploit?
2. *Ecological rationality* What environmental structures can a given heuristic exploit, that is, where is it successful and where will it fail? A heuristic is not good or bad, rational or irrational, per se, but only relative to environmental structures.
3. *Design of heuristics and environments* How can heuristics be designed to solve a given problem? How can environments be designed to support the mind in solving a problem?

The first question is descriptive, concerning the content of *Homo sapiens'* adaptive toolbox. The tools in the toolbox are the heuristics, and the term "adaptive" refers to the well-documented fact that people tend to adjust the heuristics they use to the environment or problem they encounter. The second question is normative. The rationality of a heuristic is not logical, but ecological—it is conditional on environmental structure. The study of ecological rationality has produced results that appear logically impossible or counterintuitive, such as when a judgment based on only one reason is as good as or better than one based on more reasons or when partial ignorance leads to more accurate inferences about the

world than more knowledge does (Gigerenzer, 2004). For instance, environmental structures such as high predictive uncertainty, small samples, and skewed cue validities allow the simple "take the best" heuristic, which ignores most information, to make more accurate predictions than do multiple regression or neural networks that integrate all information and use sophisticated calculation (Brighton, 2006; Chater, Oaksford, Nakisa, & Redington, 2003; Martignon & Hoffrage, 2002). Less can be more. The third question concerns cognitive (environmental) engineering. It draws on the results of the study of ecological rationality to design heuristics for given problems, such as whether or not a child should be given antibiotics (Fischer et al., 2002), or to design environments so that they fit the human mind, such as determining how to represent DNA evidence in court so that judges and jurors understand it (Hoffrage, Hertwig, & Gigerenzer, 2000).

Heuristics are embedded in (social) environments. For the reserve policemen, the environment included Major Trapp and the other men; the organ donors' environment is shaped by the legal default. Their actions are explained by both heuristics *and* their respective environments. This type of explanation goes beyond accounts of moral action in terms of personality traits such as an authoritarian personality, attitudes such as anti-Semitism, or prejudices against minorities or majorities. Unlike traits, attitudes, and preferences, which are assumed to be fairly stable across situations, heuristics tend to be highly context sensitive (Payne et al., 1993). A single policeman isolated from his comrades might not have hesitated to step forward.

If moral action is based on fast and frugal heuristics, it may conflict with traditional standards of morality and justice. Heuristics seem to have little in common with consequentialist views that assume that people (should) make an exhaustive analysis of the consequences of each action, nor with the striving for purity of heart that Kant considered to be an absolute obligation of humans. And they do not easily fit a neo-Aristotelian theory of virtue or Kohlberg's sophisticated postconventional moral reasoning. The closest cousin within moral philosophy seems to be rule utilitarianism (rather than act utilitarianism), which views a particular action as being right if it is consistent with some moral rule, such as "keep promises" (Downie, 1991). As mentioned before, heuristics provide explanations of actual behavior; they are not normative ideals. Their existence, however, poses normative questions.

What can be gained from analyzing moral actions in terms of fast and frugal heuristics? I believe that there are two goals:

1. *Explanation of moral actions* The first result would be a theory that explains the heuristic processes underlying moral actions, just as for judgment and decision making in general. Such a theory is descriptive, not normative.
2. *Modification of moral actions* The adaptive nature of heuristics implies that moral actions can be changed from outside, not just from inside the mind. Changes in environments, such as institutions and representations, can be sufficient to foster desired behavior and reduce moral disaster.

To illustrate the second goal, consider again the case of organ donation. A legal system aware of the fact that heuristics rather than reasoned preferences tend to guide behavior can make the desired option the default. In the United States, simply switching the default would save the lives of many patients who otherwise wait in vain for a donor. At the same time, this measure would save the expenses of current and future donor campaigns, which are grounded on an inadequate theory of mind. Setting proper defaults provides a simple solution for what looks like a complex moral problem. Similarly, consider once again the men of Reserve Police Battalion 101. With his offer, Major Trapp brought the Judaeo-Christian commandment "Don't murder," with which the Hamburg men grew up, into conflict with the "Don't break ranks" heuristic. With knowledge of the heuristic guiding his men's behavior, Major Trapp could have made a difference. He could have framed his offer the other way around, so that not breaking ranks no longer conflicted with not killing. Had he asked those who *felt up to the task* to step out, the number of men who participated in the killing might have been considerably smaller. This cannot be proven; yet, like Browning, I suspect that situational factors *can* shape moral behavior, as the prison experiments by Philip Zimbardo and the obedience experiments by Stanley Milgram indicate. These cases exemplify how a theory of heuristics could lead to instructions on how to influence moral action "from outside."

What are the limits of the heuristics approach? I do not believe that my analysis promises a normative theory of moral behavior. Yet the present descriptive analysis can put constraints on normative theories. A normative theory that is uninformed as to the workings of the mind, or is impossible to implement in a mind (or machine), will most likely not be useful for making our world better (see below).

Embodiment and Situatedness

Heuristics allow us to act fast—a requirement in situations where deferring decisions until more information is available can do harm to a

person, such as in emergency unit decisions. Heuristics are frugal, that is, they ignore part of the information, even when it is available. Finally, heuristics can perform well because they are embodied and situated. Let me illustrate these features by an example that has nothing to do with moral action.

How does a player catch a fly ball? If you follow a classical information-processing approach in cognitive science, you assume that the player needs a more or less *complete representation* of the environment and a sophisticated computer to calculate the trajectory from this representation. To obtain a complete representation, the player would have to estimate the ball's initial velocity, angle, and distance, taking account of air resistance, wind speed, direction of wind, and spin. The player would then calculate the trajectory and run to the point where the ball will hit the ground. All this creates a nice optimization model, but there is no empirical evidence for it. No mind or machine can solve the problem this way. In the real world, players do not compute trajectories; instead, they rely on a number of simple heuristics. One is the *gaze heuristic*, which works if the ball is already high up in the air:

Fixate your gaze on the ball, start running, and adjust your speed so that the angle of gaze remains constant.

The gaze heuristic *ignores* all causal information necessary to compute the trajectory. It does not need a complete representation, even if it could be obtained. The heuristic uses only one piece of information, the angle of gaze. Yet it leads the player to the point where the ball will land. If you ask players how they catch a ball, most do not know the heuristic or can describe only one building block, such as "I keep my eye on the ball." The heuristic is composed of building blocks that draw on specific abilities. "Fixate your gaze on the ball" is one building block of the heuristic, which exploits the evolved ability to track a moving object against a noisy background. In general, a fast and frugal heuristic is a rule that is anchored in both mind and environment:

1. *Embodiment* Heuristics exploit evolved abilities, such as the human ability for group identification, imitation, or cheating detection (e.g., Cosmides & Tooby, 2004). The gaze heuristic exploits the ability of object tracking, that is, the ability to track a moving target against a noisy background, which emerges in three-month-old infants (Rosander & Hofsten, 2002). The default heuristic exploits a set of evolved abilities that deal with cooperation in small groups of people, such as imitation and trust.

2. *Situatedness* Heuristics exploit environmental structures, such as social institutions or the redundancy of information. The gaze heuristic even manipulates the environment, that is, it transforms the complex relation between player and ball into a simple, linear one.

Evolved abilities allow heuristics to be simple. Today's robots cannot trace moving objects against noisy backgrounds as well as humans; thus, the gaze heuristic is only simple for the evolved brains of humans, fish, flies, and other animals using it for predation and pursuit. The embodiment of heuristics poses a problem for the view that mental software is largely independent of the hardware and that mental processes can be realized in quite different physical systems. For instance, Hilary Putnam (1960) used Alan Turing's work as a starting point to argue for a distinction between the mind and the brain in terms of the separation of software from hardware. For many psychologists, this seemed a good basis for the autonomy of psychology in relation to neurophysiology. The rhetoric was that of cognitive systems that describe the thought processes "of everything from man to mouse to microchip" (Holland, Holyoak, Nisbett, & Thagard, 1986, p. 2). In contrast, heuristics do not function independently of the brain; they exploit it. Therefore, the heuristics used by "man and microchip" should not be the same. In summary, heuristics are simple because they exploit human brains—including their evolved abilities. This position is inconsistent with the materialistic ideal of reducing the mind to the brain, and also with the dualistic ideal of analyzing the mind independent of the brain, and vice versa.

Environmental structures allow heuristics to function well. When a clear criterion of success exists, one can mathematically analyze in which environments a given heuristic will succeed or fail. For instance, the gaze heuristic only works well when the ball is already high up in the air, not beforehand. In the latter case, the third building block of the heuristic needs to be changed into "adjust your speed so that the image of the ball is rising at a constant speed" (Shaffer, Krauchunas, Eddy, & McBeath, 2004). This illustrates that one does not need to develop a new heuristic from scratch for every new situation but can perhaps just modify one building block. The analysis of the situations in which a given heuristic works and fails is called the study of its "ecological rationality." The study of ecological rationality is difficult to generalize to moral action, unless criteria for success are supplied. Such criteria need to be precise; vague notions such as happiness and pleasure are insufficient for a mathematical analysis of ecological rationality.

Moral Action and Heuristics

I propose three hypotheses. First, moral intuitions as described in the social intuitionist theory (e.g., Haidt, 2001) can be explicated in terms of fast and frugal heuristics (Gigerenzer, 2007). Let me elaborate with a frequently posed distinction: Is moral judgment based on reasons or feelings? According to the philosophical theory of intuitionism, "a person who can grasp the truth of true ethical generalizations does not accept them as the result of a process of ratiocination; he just sees without argument that they are and must be true, and true of all possible worlds" (Harrison, 1967, p. 72). This view makes strong assumptions (that ethical generalizations are synthetic and a priori) and is hard to refute, as Harrison describes in detail. However, the idea that moral judgments are caused by perception-like, self-evident moral intuitions (not necessarily moral truths) has become the fundament of the social intuitionist approach to moral judgment. In this view, "moral reasoning does not cause moral judgment; rather moral reasoning is usually a post hoc construction, generated after a judgment has been reached" (Haidt, 2001, p. 814). Just like its philosophical sibling, social intuitionist theory makes a descriptive claim, and the evidence presented includes the sudden appearance in consciousness of moral judgments, after which people are "morally dumbfounded," that is, they mostly cannot tell how they reached a judgment (Haidt, Algoe, Meijer, Tam, & Chandler, 2000; Nisbett & Wilson, 1977). The unresolved issue in this theory is that "moral intuition" remains an unexplained primitive term.[2]

I agree with the proposition that in many cases moral judgments and actions are due to intuitive rather than deliberative reasoning. I also grant that there are important exceptions to this hypothesis, such as Benjamin Franklin's (1772/1987) "moral algebra" and the professional reasoning of judges. However, reasons given in public can be post hoc justification. What intuitionist theories could gain from the science of heuristics is to explicate intuition in terms of fast and frugal heuristics. This would provide an understanding of how intuitions are formed.

Here is my second hypothesis: Heuristics that underlie moral actions are largely the same as those for underlying behavior that is not morally tinged. They are constructed from the same building blocks in the adaptive toolbox. That is, one and the same heuristic can solve both problems that we call moral and those we do not. For instance, the "do what the majority do" heuristic (Laland, 2001) guides behavior in a wide range of situations, only some of which concern moral issues:

If you see the majority of your peers behave in a certain way, engage in the same action.

This heuristic produces social facilitation and guides behavior through all states of development from childhood to teenage and adult life. It virtually guarantees social acceptance in one's peer group. It can steer consumer behavior (what clothes to wear, what CDs to buy) and moral action as well (to donate to a charity, to discriminate against minorities). Teenagers tend to buy Nike shoes because their peers do, and skinheads hate foreigners for no other reason than that their peers hate them as well. The second hypothesis implies that moral intuitions are based on reasons, just as in cognitive heuristics, thus questioning the original distinction made between feelings and reasons. By explicating the processes underlying "feeling" or "intuition," the feeling/reason distinction is replaced by one between the conscious versus unconscious reasons that cause moral judgments.

The third hypothesis is that the heuristics underlying moral action are generally unconscious. If one interviews people, the far majority are unaware of their underlying motives. Rather, they often stutter, laugh, and express surprise at their inability to find supporting reasons for their likes and dislikes, or they invent post hoc justifications (Haidt, 2001; Haidt & Hersh, 2001; Nisbett & Wilson, 1977; Tetlock, 2003). This lack of awareness is similar to decision making outside the moral domain. As mentioned before, baseball players are often unaware of the heuristics they use, and consumers are not always able to explain why they bought a particular car, dress, or CD. Because of their simplicity and transparency, however, heuristics can be easily made conscious, and people can learn to use or to avoid them.

The view that moral action is based on fast and frugal heuristics also has three methodological implications:

1. *Study social groups in addition to isolated individuals* Heuristics exploit evolved abilities and social motives, such as the human potential for imitation, social learning, and feelings of guilt (Gigerenzer & Hug, 1992). The methodological implication is to study behavior in situations where these heuristics can unfold, such as in the presence of peers (e.g., Asch's [1956] conformity experiments). Compare the situation that the men of Reserve Police Battalion 101 faced with the hypothetical moral dilemmas in which an individual has to choose either to kill one person or otherwise let twenty people be killed by someone else (e.g., Williams, 1988). Here, the

experimental participant is studied in isolation. Heuristics such as "*don't break ranks*" and "*do what the majority do*" can hardly be detected.

2. *Study natural environments in addition to hypothetical problems* The science of heuristics aims for theoretical statements that involve the pairing of heuristics with environments, where the environment may select a heuristic or the heuristic may shape the environment (Gigerenzer et al., 1999). The methodological implication is to study moral intuitions in natural environments, or in experimental models thereof (e.g., Zimbardo's prison experiments and Milgram's obedience studies) rather than using hypothetical problems only. Toy problems such as the "trolley problems" eliminate characteristic features of natural environments, such as uncertainty about the full set of possible actions and their consequences, and do not allow the search for more information and alternative courses of action. I am not suggesting that hypothetical moral problems are of no use but that the present focus on hypothetical problems in experimental moral psychology as well as in moral philosophy creates a limited opportunity for understanding moral action. Because heuristics used tend to be very sensitive to social context, the careful analysis of natural environments is essential. This focus on the environment contrasts with those cognitive theories that assume, implicitly or explicitly, that morality is located within the individual mind, like a trait or a set of knowledge structures. For instance, in Kohlberg's (1971) rational cognitive theory, inspired by Piaget's (1932/1965) step model, moral development is a process that can be fully described internally, from egoistic to conventional to postconventional forms of reasoning. In these internalistic views, the structure of the environment appears of little relevance.

3. *Analyze moral behavior in addition to self-reports* People are typically unaware of the heuristics underlying their moral judgments or understand only part of them. The methodological implication is that asking people for reasons will rarely reveal the heuristics on which they actually base their decisions. Observation and analysis of behavior are indispensable if one wants to understand what drives people.

I will illustrate these points with judgments of trustworthiness in the legal context. The results of the following case study indicate that (1) legal decision makers use fast and frugal heuristics, (2) their heuristics have the same structure (not content) as heuristics used to solve nonmoral problems, (3) magistrates are largely unaware of this fact and believe their decisions are based on elaborate reasoning, and (4) the heuristics appear to be shaped by the social institution in which the decision makers operate.

Bail Decisions and Due Process

One of the initial decisions of the legal system is whether to bail the defendant unconditionally or to make a punitive decision such as custody or imprisonment. The bail decision is not concerned with the defendant's guilt but with his or her moral trustworthiness: whether or not the defendant will turn up at the court hearing, try to silence witnesses, or commit another crime. In the English system, magistrates are responsible for making this decision. About 99.9% of English magistrates are members of the local community without legal training. The system is based on the ideal that local justice be served by local people.

In England and Wales, magistrates make decisions on some two million defendants per year. They sit in court for a morning or afternoon every one or two weeks and make bail decisions as a bench of two or three. The Bail Act of 1976 and its subsequent revisions (Dhami & Ayton, 2001) require that magistrates pay regard to the nature and seriousness of the offense; to the character, community ties, and bail record of the defendant; and to the strength of the prosecution case, the likely sentence if convicted, and any other factor that appears to be relevant. Yet the law is silent on how magistrates should weigh and integrate these pieces of information, and the legal institutions do not provide feedback on whether their decisions were in fact appropriate or not. The magistrates are left to their own intuitions.

How do magistrates actually make these millions of decisions? To answer this question, several hundred trials were observed in two London courts over a four-month period (Dhami, 2003). The average time a bench spent with each case was less than 10 minutes. The analysis of the actual bail decisions indicated a fast and frugal heuristic that accounts for 95% of all bail decisions in Court A (see figure 1.1, left; cross-validation performance: 92%). When the prosecution requested conditional bail, the magistrates also made a punitive decision. If not, or if no information was available, a second reason came into play. If a previous court had imposed conditions or remanded in custody, then the magistrates also made a punitive decision. If not, or if no information was available, they followed the action of the police.

The bail decisions in Court B could be modeled by the same heuristic, except that one of the reasons was different (see figure 1.1, right). The benches in both courts relied on the same defensive rationale, which is known as "passing the buck." The magistrates' heuristics raise an ethical issue. In both London courts, they violate due process. Each bench based

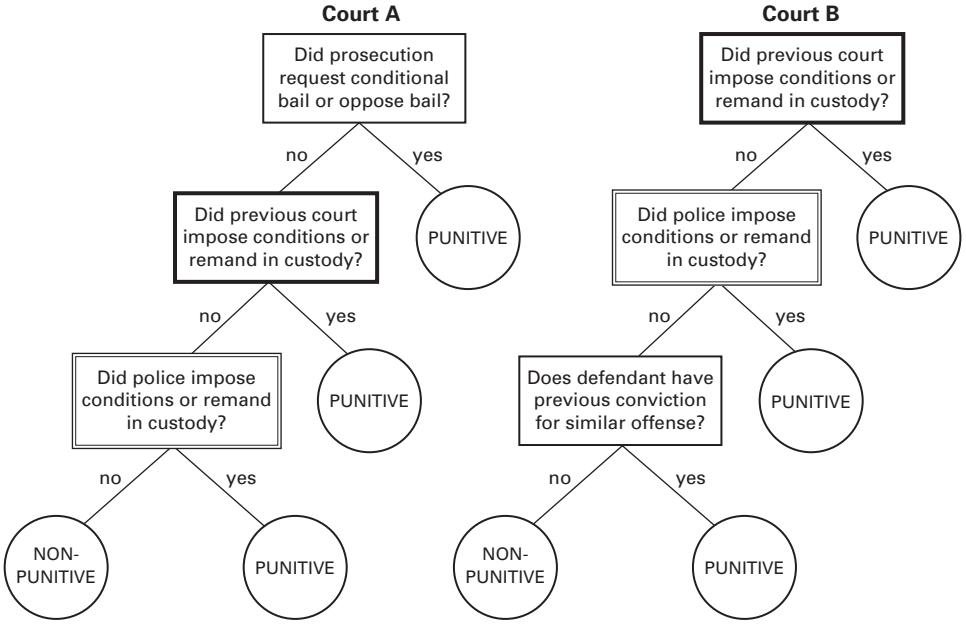

Figure 1.1
Models of fast and frugal heuristics for bail decisions in two London courts (adapted from Dhami, 2003).

a punitive decision on one reason only, such as whether the police had imposed conditions or imprisonment. One could argue that the police or prosecution had already looked at all the evidence concerning the defendant, and therefore magistrates simply used their recommendation as a shortcut (although this argument would make magistrates dispensable). However, the reasons guiding the heuristics were *not* correlated with the nature and seriousness of the offense or with other pieces of information relevant for due process.

The bail study investigated magistrates in their original social context (a bench of two or three laypeople) and in their natural environment (magistrates work in an institution that provides no systematic feedback about the quality of their decisions, and they can only be proven wrong if they bailed a defendant who then committed an offense; see below). Are its results consistent with the three hypotheses? With respect to the first hypothesis, the bail study can, at best, provide proof of the existence of fast and frugal heuristics but does not allow the conclusion that a substantial part of moral action is based on them. The answer to the second

hypothesis, that the structure of moral heuristics mirrors that of other heuristics, however, is positive. The two bail heuristics have the same structure as a class of cognitive heuristics called "fast and frugal trees" (Katsikopoulos & Martignon, 2004). Unlike in a full tree, a decision is possible at each node of the tree. For three binary reasons with values [0, 1], where "1" allows for an immediate decision, the general structure of a fast and frugal tree is as follows:

Consider the first reason: If the value is "1," stop search and choose the corresponding action. Otherwise,
Consider the second reason. If the value is "1," stop search and choose the corresponding action. Otherwise,
Consider the third reason: If the value is "1," choose action A; otherwise choose B.

Fast and frugal trees are a subclass of heuristics that employ sequential search through reasons (Gigerenzer, 2004). The bail heuristics embody a form of one-reason decision making: Although more than one reason may be considered, the punitive decision itself is based on only one reason. The decision is noncompensatory, which means that reasons located further down the tree cannot compensate for or overturn a decision made higher up in the tree. In other words, the heuristic makes no trade-offs. Note that sequential heuristics can embody interactions, such as that bail is given only if neither prosecution, nor previous court, nor police opposed bail. Fast and frugal trees play a role in situations beyond the trustworthiness of a defendant, such as in medical decision making (Fischer et al., 2002; Green & Mehr, 1997).

Third, are magistrates aware of what underlies their judgments? When asked to explain their decisions, their stories were strikingly different. A typical answer was that they thoroughly examined all the evidence on a defendant in order to treat the individual fairly and without bias, and that they based their decision on the full evidence. For instance, one explained that the decision "depends on an enormous weight of balancing information, together with our experience and training" (Dhami & Ayton, 2001, p. 163). Another said that "the decisions of magistrates are indeed complex, each case is an 'individual case'" (Dhami, 2001, p. 255). Furthermore, magistrates actually asked for information concerning the defendant, which they subsequently ignored in their decisions. Unless the magistrates deliberately deceived the public about how they make bail decisions (and I have no grounds to assume so), one must conclude on the basis of the models in figure 1.1 that they are largely unaware of the

heuristics they use. This dissociation between the reported reasons and the actual reasons (as modeled in the bail heuristics) is consistent with what Konecni and Ebbesen (1984) refer to as the "mythology of legal decision making" (p. 5).

Models of Moral Heuristics

There is a classical distinction between rationalist and nonrationalist theories of moral judgment. Is moral judgment the result of reasoning and reflection, as in Kohlberg's (1969) and Piaget's (1932/1965) theories? Or is it an intuitive process, as in Haidt's (2001) social intuitionism perspective, based on Hume's ideas? Rationalist theories assume that reasoning comes first and that moral intuition is its product, whereas social intuitionist theories assume that moral intuition typically comes first and reasoning is a post hoc attempt to justify an intuition to an audience. I suggest that the intuitions can be explicated by heuristics relying on reasons. The opposition is not between intuition and reasoning, in my view, but between the (unconscious) reasons underlying intuition and the conscious, after-the-fact reasons. The magistrates' judgments, for instance, can be explained by a simple heuristic based on three reasons, yet they believed they were engaging in highly complex reasoning. This point fits well with the social intuitionist view of moral judgment, where rationalization is ex post facto rather than the cause of the decision (Haidt, 2001). Moreover, the heuristics perspective can extend the intuitionist view in two directions: It provides an analysis of the heuristic process and of the environment.

Why Processes Models Are Essential

Unlike views that treat intuition as an unexplained primitive notion or attribute it to feelings as opposed to reasons, the heuristics perspective asks to specify models of what underlies moral intuition. The descriptive goal of the heuristics program is to spell out what the heuristics underlying intuition are and how they differ from the post hoc rationalization of one's judgment. This is a call for *models* and for going beyond mere *labels* for heuristics, such as "availability" and "representativeness" (Kahneman & Tversky, 1996). Mere labels and ying–yang lists of dichotomies such as "System 1 versus System 2" can account post hoc for everything and nothing (Gigerenzer, 1996, 1998; Gigerenzer & Regier, 1996). For decades, these surrogates for theories have hindered progress in the psychology of judgment. We instead need testable theories of cognitive processes, such as shown in figure 1.1. It is a striking paradox that many cognitive and

social psychologists practice "black-box behaviorism." They don't seem to dare or care to open the box more than an inch, and they throw in a "one-word explanation" (e.g., salience, availability) before quickly shutting it again. B. F. Skinner would have been happy to see cognitive psychologists voluntarily abstain from theories of cognitive processes.

Models of heuristics demonstrate that the dichotomy between intuitions and reasons has its limits. Like conscious reasoning, sequential search heuristics—as shown in figure 1.1—rely on reasons. After all, one could make the bail heuristics public, implement them into a computer program, and replace the entire British bail system. Moral intuitions can be based on reasons, even if the latter are unconscious. These reasons, however, need not be the same as those given post hoc in public. In addition, moral intuition can ignore most or even all reasons, as in the case of simply copying the moral action of one's peers.

Yet why do we need models of heuristic processes underlying moral intuitions? Could one not simply say that people behave *as if* they were maximizing justice, well-being, or happiness? Consider the task of catching a ball again. Could one not simply say, as the biologist Richard Dawkins (1989) put it, "When a man throws a ball high in the air and catches it again, he behaves as if he had solved a set of differential equations in predicting the trajectory of the ball" (p. 96)? As-if theories do not describe how people actually solve a problem, in courts or sports. However, not knowing the heuristics can have unwanted consequences. I once gave a talk on the gaze heuristic, and a professor of business administration came up to me and told me the following story. Phil (not his real name) played baseball for the local team. His coach scolded him for being lazy, because Phil sometimes trotted over, as others did, toward the point where the ball came down. The angry coach insisted that he instead run as fast as he could. However, when Phil and his teammates tried to run at top speed, they often missed the ball. Phil had played as an outfielder for years and had never understood how he caught the ball. Unaware of the gaze heuristic and the other heuristics players use, the coach assumed something like the as-if model and did not realize that the heuristic dictates the speed at which a player runs, and that running too fast will impede performance. Phil's case illustrates that knowing the heuristic can be essential to correcting wrong conclusions drawn from an as-if model.

I argue that in the moral domain it is equally important to analyze the processes underlying people's helpful or harmful behavior in order to improve a situation. For instance, by starting with the assumption that the magistrates behaved as if they were maximizing the welfare of defendants

and society, one would miss how the system works and not be able to improve it. Now that we have—for the first time—a good model of the magistrates' underlying heuristics, it is possible to assess the system and ask the critical questions. Are magistrates necessary at all? And, if the answer is positive, how can one improve their heuristics as well as a legal system that supports defensive justice rather than due process?

Institutions Shape Intuitions

The science of heuristics emphasizes the analysis of the "external" environment, in addition to the "internal" heuristics. Heuristics that shape moral intuitions are in part a consequence of the external environment, and vice versa. How does an institution shape heuristics?

The legal institution in which magistrates operate seems to support their mental dissociation. The law requests that magistrates follow due process. The magistrates' official task is to do justice to a defendant and the public, that is, to minimize the two possible errors one can make. This first error occurs when a suspect is released on bail and subsequently commits another crime, threatens a witness, or does not appear in court. The second error occurs when a suspect who would not have committed any of these offenses is imprisoned. However, as mentioned before, English legal institutions collect no systematic information about the quality of magistrates' decisions. Even if statistics were kept about when and how often the first error occurs, it would be impossible to do the same for the second error, simply because one cannot find out whether an imprisoned person would have committed a crime if he or she had been bailed. That is, the magistrates operate in an institution that does not or cannot provide feedback about how well they protect the defendant and the public. They effectively cannot learn how to solve the intended task, and the bail heuristics suggest that they instead try to solve a different one: to protect themselves rather than the defendant. Magistrates can only be proven to have made a bad decision if a suspect who was released committed an offense or crime while on bail. If this happens, the bail heuristic protects them against accusations by the media or the victims. The magistrates in Court A, for instance, can always argue that neither the prosecution, nor a previous court, nor the police had imposed or requested a punitive decision. Thus, the event was not foreseeable. An analysis of the institution can help to understand the nature of the heuristics people use and why they believe they are doing something else.

More generally, consider an institution that requires their employees to perform a duty. The employees can commit two kinds of errors: false alarms

and misses. If an institution (1) does not provide systematic feedback concerning false alarms and misses but (2) blames the employees if a miss occurs, the institution fosters employees' self-protection over the protection of their clients, and it supports self-deception. I call this environmental structure a "split-brain institution." The term is borrowed from the fascinating studies of people whose corpus callosum—the connection between the right and left cerebral hemispheres—has been severed (Gazzaniga, 1985). Split-brain patients confabulate post hoc stories with the left (verbal) side of their brain to rationalize information or phenomena perceived by the right (nonverbal) side of their brain, which they are apparently unaware of. The analogy only holds to a point. Unlike a split-brain patient, a split-brain institution can impose moral sanctions for confabulating and punishment for awareness of what one does. If magistrates were fully aware of their heuristics, a conflict with the ideal of due process would ensue. Medical institutions often have a similar split-brain structure. Consider a health system that allows patients to visit a sequence of specialized doctors but does not provide systematic feedback to these doctors concerning the efficacy of their treatments, and in which doctors are likely to be sued by the patient for having overlooked a disease but not for overtreatment and overmedication. Such a system fosters doctors' self-protection over the protection of their patients and supports similar self-deception as in the case of the magistrates.

Should We Rely on Moral Heuristics?

The answer seems to be "no." Heuristics ignore information, do not explore all possible actions and their consequences, and do not try to optimize and find the best solution. Thus, for those theories that assume that all consequences of all possible actions should be taken into account to determine the best action, fast and frugal heuristics appear to be questionable guidelines. Even social intuitionists who argue against rationalist theories as a valid descriptive theory are anxious not to extend their theory to the normative level. For instance, Haidt (2001) is quick to point out that intuition is not about how judgments should be made, and he cites demonstrations that "moral intuitions often bring about nonoptimal or even disastrous consequences in matters of public policy, public health, and the tort system" (p. 815). Understanding nonrational intuitions may 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" (p. 815). The same negative conclusion can be derived from

the heuristics-and-biases program (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 2000), where heuristics are opposed to the laws of logic, probability, or utility maximization, and only the latter are defended as normative. Sunstein (2005), for instance, applies this approach to moral intuitions and emphasizes that heuristics lead to mistaken and even absurd moral judgments. Just as Kahneman and Tversky claimed to know the only correct answer to a reasoning problem (a controversial claim; see Gigerenzer 1996, 2000), Sunstein has a clear idea of what is right and wrong for many moral issue he discusses, and he holds people's moral heuristics responsible for their negligence, wrong-doing, and evil. Despite laudable attempts to propose models of heuristics, he relies on vague terms such as "availability" and "dual-process models." Yet, without some degree of precision, one cannot spell out in what environment a given heuristic would work or not work. All these views seem to converge to a unanimous consensus: Heuristics are always second-best solutions, which describe what people do but do not qualify as guidelines for moral action.

The view that heuristics can be prescriptive, not only descriptive, distinguishes the study of the adaptive toolbox from the heuristics-and-biases program. Both programs reject rational choice or, more generally, optimization as a general descriptive theory, arguing instead that people often use heuristics to make decisions. However, the heuristics-and-biases program stops short when it comes to the question of "ought." The study of ecological rationality, in contrast, offers a more radical revision of rational choice, including its prescriptive part. The gaze heuristic illustrates that ignoring all causal variables and relying on one-reason decision making can be ecologically rational for a class of problems that involve the interception of moving objects. More generally, today we know of conditions under which less information is better than more, for instance, when relying on only one reason leads to predictions that are as good as or better than by weighting and adding a dozen reasons (Czerlinski, Gigerenzer, & Goldstein, 1999; Hogarth & Karelaia, 2005; Martignon & Hoffrage, 1999). We also understand situations where limited cognitive capacities can enable language learning (Ellman, 1993) and covariation detection (Kareev, 2000) better than larger capacities do (Hertwig & Todd, 2003). Simple heuristics, which ignore part of the available information, are not only faster and cheaper but also more accurate for environments that can be specified precisely. I cannot go into detail here, but the general reason for these counterintuitive results are that, unlike logic and rational choice models, heuristics exploit evolved abilities and structures of environments, including their uncertainty (Gigerenzer, 2004). This opens up the

possibility that when it comes to issues of justice and morals, there are situations in which the use of heuristics, as opposed to an exhaustive analysis of possible actions and consequences, is preferable.

Can heuristics be prescriptive? As I said earlier, unlike in inferences and predictions where a clear-cut criterion exists, in the moral domain one can only analyze the situations in which a heuristic is ecologically rational if a normative criterion is introduced. For inference tasks, such as classification and paired comparison, heuristics are evaluated by predictive accuracy, frugality, speed, and transparency. Some of these criteria may be relevant for moral issues. For instance, transparent rules and laws may be seen as a necessary (albeit not sufficient) condition for creating trust and reassurance in a society, whereas nontransparent rules and arbitrary punishments are the hallmark of totalitarian systems. Transparency also implies that the number of laws is few, as in the Ten Commandments of Christianity. In many situations, however, there is no single agreed norm. But I believe that one should face rather than deny normative uncertainty.

Last but not least, the science of heuristics can provide a better understanding of the limits of normative theories of morality. I illustrate this point with versions of consequentialism and similar theories that are based on the moral ideal of maximization.

The Problem with Maximization Theories

The idea that rational choice means the maximization of the expected value has been attributed to the seventeenth-century French mathematicians Blaise Pascal and Pierre Fermat and dated to their exchange of letters in 1654. Pascal used the calculus for a moral problem: whether or not to believe in God (Daston, 1988). He argued that this decision should not be based on blind faith or blind atheism but on considering the consequences of each action. There are two possible errors. If one believes in God but he does not exist, one will forgo a few worldly pleasures. However, if one does not believe in God but he exists, eternal damnation and suffering will result. Therefore, Pascal argued, however small the probability that God exists, the known consequences dictate that believing in God is rational. What counts are the consequences of actions, not the actions themselves. "Seek the greatest happiness of the greatest number"—the slogan associated with Jeremy Bentham—is a version of this maximization principle, a form of hedonistic utilitarianism where the standard is not the agent's own happiness but that of the greatest number of people. Today, many forms of utilitarianism and consequentialism exist, both normative and descriptive (see Smart, 1967).

My question is, can utilitarianism and consequentialism provide (1) a norm and (2) a description of moral action in the real world? I emphasize the "real world" as opposed to a textbook problem such as the trolley problem, where I assume, for the sake of argument, that the answer is "yes." When I use the term *consequentialism* in the following, I refer to theories that postulate maximizing, implicitly or explicitly: "in any form of direct consequentialism, and certainly in act-utilitarianism, the notion of the right action in given circumstances is a maximizing notion" (Williams, 1988, p. 23; see also Smart, 1973). Just as in Pascal's moral calculus, and Daniel Bernoulli's maximization of subjective expected utility, this form of consequentialism is about the *optimal* (best) action, not just one that is good enough. It demands optimizing, not satisficing.

To find the action with the best consequences is not a simple feat in the real world. It requires determining the set of all possible actions, all possible consequences, their probabilities, and their utilities. There are at least four interpretations of this process of maximizing:

- a conscious mental process (e.g., to think through all possible actions and consequences),
- an unconscious mental process (the brain does it for you, but you don't notice the process, only the result),
- an *as-if* theory of behavior (people behave *as if* they computed the action with the highest utility; no claim for a conscious or unconscious mental process), and
- a normative goal (maximizing determines which action one ought to choose; no claim as a model of either process or behavior).

Proponents of consequentialism have emphasized that their theory is not just a fiction created by some philosophers and economists and handed down to moral scholars. It is written into law (Posner, 1972). According to U.S. tort law, an action is called "negligent" and the actor is likely to pay damages if the probability of resulting harm multiplied by the cost of harm to others exceeds the benefit of the action to the actor. This is known as the "Learned Hand formula," named after Judge Learned Hand, who proposed the formula for determining negligence in 1947. If the expected damage is less than the expected benefit, the actor is not liable for damages, even if the risked harm came to pass.

Yet there is a second part to this story. Although Judge Learned Hand has been acclaimed as an early proponent of maximization and consequentialism, he also held the view that "all such attempts [to quantify the determinants of liability] are illusory; and, if serviceable at all, are so only to center attention upon which one of the factors may be determinate in

any given situation" (Moisan v. Loftus, 178 F.2d 148, 149 [2d Cir 1949]; see Kysar et al., 2006). Here, Judge Hand seems to favor one-reason decision making, such as fast and frugal decision trees and Take The Best (Gigerenzer & Goldstein, 1996). This second part illustrates some of the problems with maximization in the real world, which I will now turn to.

Computational Limits

By definition, consequentialism (in the sense of maximization) can only give guidelines for moral action *if the best action can actually be determined by a mind or machine*. I argue that this is typically not the case in the real world. To the degree that this is true, consequentialism is confined to well-defined moral problems with a limited time horizon and a small set of possible actions and consequences that do not allow uncertainty and surprises. Moral textbook problems in the philosophical literature have this impoverished structure (Williams, 1988).

Consider, in contrast, a moral game that has the structure of chess—a choice set of about 30 possible actions per move, a temporal horizon of a sequence of some 20 moves, and two players. One player acts, the other reacts, and so on, with 10 moves for each person. For each move, both players can choose to act in 1 out of the 30 possible ways—to tell the truth, to lie, to cheat, to form an alliance, to withhold information, to insult, to threaten, to blackmail, and so on. The opponent then responds with 1 of the 30 possible actions, and so on. The game is well defined; no negotiation of rules or self-serving changes are allowed. Every action in this game of social chess depends on those of the other person, so one has to look ahead to understand the consequences of one's own action. Can one determine the best sequence of actions in this game? Although the game looks like a somewhat impoverished human interaction, no mind or machine can enumerate and evaluate all consequences. A simple calculation will illustrate this.

For each action, there are 30 possibilities, which makes in 20 moves 30^{20} sequences, which amounts to some

350,000,000,000,000,000,000,000,000,000

possible sequences of moves. Can a human mind evaluate all of these consequences? No. Can our fastest computers do it? Deep Blue, the IBM chess computer, can examine some 200 million possible moves per second. How long would it take Deep Blue to think through all consequences in social chess and choose the move that maximizes utility? Even at its breathtaking speed, Deep Blue would need some 55,000 billion years to think 20

moves ahead and pick the best one. (Recall that the Big Bang is estimated to have occurred only some 14 billion years ago.) But 20 moves are not yet a complete game of chess or of human interaction. In technical terms, social chess is "computationally intractable" for minds and machines.

If we increased the number of people interacting from two to three or more, we would encounter a new obstacle for maximization. Just as the predictive power of physics ends with a three-body problem—such as earth, moon, and sun, moving under no influence other than their mutual gravitation—there is no best way to predict the dynamics of the mutual attractions of three or more people. This computational problem arises both in competitive and cooperative games if they are played in an uncertain world rather than in a small closed one. My conclusion is that consequentialism, understood as the maximization of some measure of utility or happiness, can only work with a limited time perspective and limited interactions. Beyond these limits, consequentialism can neither be prescriptive nor descriptive.

When Maximization Is Out of Reach

More generally, situations for which maximization—in consequentialism or other moral theories—is impossible, include the following:

1. *Computationally intractable problems* These are well-defined problems, such as chess and the computer games Tetris and Minesweeper. No mind or machine can compute the optimal solution in real time. For instance, when former chess world champion Kasparov played against the IBM chess program Deep Blue, both had to rely on heuristics. The reason is not simply because people or computers have limited cognitive capacities but because the problem is computationally intractable. This is not necessarily bad news. Games where we know the optimal solution (such as tic-tac-toe) are boring for exactly this reason. The same holds for moral issues. If social chess were computable, part of our emotional life would become obsolete. We would always know how to behave optimally, as would our partners. There would be fewer things to hope for, and less surprise, joy, disappointment, and regret.

2. *The criterion cannot be measured with sufficient precision* For instance, there is no way to optimize the acoustics of a concert hall because experts consistently disagree about what constitutes good acoustics. The same applies to many moral criteria. The criterion of consequentialist theory—"happiness," "pleasure," or "utility"—is at least as difficult to measure as the acoustics of a concert hall, and for similar reasons. People, including

experts, will not agree what consequences make them and others most happy, and there are societies where happiness means little in comparison to religious faith and loyalty to one's national and cultural identity. Thus, the criterion of the greatest happiness for everyone may become increasingly fuzzy the farther one travels from one's social group. The same problem arises for norms of egalitarianism. Moral philosophers have long discussed what should be equal: opportunity, rights, income, welfare, capabilities, or something else? A heuristic may focus on those few that can be observed most easily in a given situation (Messick, 1993). However, the general problem of optimizing equality has no solution because of lack of sufficient precision.

3. *Multiple goals or criteria* Optimization is, in general, impossible for problems with multiple criteria. One cannot maximize several criteria simultaneously (unless one combines them by, say, a linear function). For instance, even if the traveling salesman problem could be solved (it cannot for large numbers of cities), its real-world equivalent has multiple criteria, not only the shortest route. These can involve the fastest route, the cheapest route, and the most scenic route. Multiple criteria or goals, however, are characteristic of moral dilemmas. Examples are paternity cases where one wants to find the truth but also protect the child from being uprooted, while doing justice to the rights of the genetic parents and the foster family. Similarly, when happiness is not of one kind, but of several, one cannot maximize all of these simultaneously.

4. *Calculative rationality can be seen as morally unacceptable* In certain domains, the idea of choosing the option with the best anticipated consequences can violate people's moral sense. These include kinship, friendship, and mate choice. When a man (or woman) proceeds rationally by empirically investigating all potential partners, the possible consequences of living with them, and the probabilities and utilities of each consequence, moral outrage from those being investigated can result. In 1611, for instance, the astronomer Johannes Kepler began a methodical search for his second wife after an arranged and unhappy first marriage. He investigated 11 possible replacements within 2 years. Friends urged him to marry Candidate No. 4, a woman of high status and tempting dowry, but she eventually rejected him for toying with her too long. The attempt to rationally determine the best alternative can be perceived as morally repulsive. Former First Lady Barbara Bush, in contrast, seemed to have undertaken little comparative study: "I married the first man I ever kissed. When I tell this to my children, they just about throw up" (Todd & Miller, 1999).

5. *Optimization can destroy trust* If an employer tried to optimize and dismissed his employees and subcontracters every year in order to hire the best ones, he might destroy loyalty, identification, and trust (Baumol, 2004). In contrast, heuristics such as satisficing entail an implicit promise to current employees that as long as their performance and development continue to be satisfactory, that is, meet an aspiration level, no changes will be made. This makes it attractive for employees to adapt their services to the needs of the firm. The value in commitments holds outside of business environments. When a university admits a graduate student, it is typically understood that there is no annual contest among students inside and outside the university for reallocating stipends competitively to the students who look best at any point in time. Rather, the support will be continued for the next several years, as long as the student's performance continues to meet standards of acceptability.
6. *Ill-defined problems* Most problems in the real world are ill-defined, that is, the set of possible actions is not known, their consequences cannot be foreseen, the probabilities and utilities are unknown, and the rules of the game are not fixed but are negotiated during the game. In these situations, maximization—of collective happiness or anything else—is, by definition, impossible.

The fact that maximization (optimization) is typically out of reach in the real world is widely ignored in philosophy, economics, and the cognitive sciences. This state of affairs has been called the "fiction of optimization" (Klein, 2001; see also Selten, 2001). Several tools for rescuing maximization are in use. One is to assume that people are unboundedly rational, that is, that they know all actions, consequences, and other information needed to calculate the best option. A second tool is to edit a computationally intractable real-world problem into one that is small enough so that the optimization calculus can be applied. However, as Herbert Simon (1955, p. 102) argued long ago, there is a complete lack of evidence that in real-world situations of any complexity, these computations can be or actually are performed. In contrast, in applied sciences such as robotics and machine learning, it is common wisdom that in order to solve real-world problems, one needs to develop heuristic methods.

Toward an Investigation of Moral Heuristics

In this essay, I argued that many—not all—moral actions can be understood as based on fast and frugal heuristics. Specifically, moral intuitions can be explicated by models of heuristics. These heuristics are strong

enough to act upon, yet people are typically not aware of their underlying rationale. Understanding heuristics requires an analysis of the social environment in which people act, because heuristics take advantage of environments and environments select heuristics. Analyzing the environment also helps to understand systematic discrepancies between the reasons people give for their moral intuitions and the underlying heuristics. To the degree that moral action is guided by heuristics, it can be influenced by changing the conditions that trigger a given heuristic. This includes the framing of an offer, as illustrated in the case of Major Trapp, and the choice of the default, as in the case of organ donation. Unlike theories that focus on traits, preferences, attitudes, and other internal constructs, the science of heuristics emphasizes the interaction between mind and social environment. Knowing the heuristics that guide people's moral actions can be of help in designing change that might otherwise be out of reach.

Notes

I thank Lael Schooler, Walter Sinnott-Armstrong, Masanori Takezawa, Rona Unrau, and the members of the LIFE Max Planck International Research School for their helpful comments.

1. Browning (1993, p. xvii). I chose this sensitive example because it is one of the best-documented mass murders in history, with the unique feature that the policemen were given the opportunity not to participate in the killing. My short account cannot do justice to the complexity of the situation, and I recommend consulting Browning's book, including the afterword, in which he deals with his critics such as Daniel Goldhagen. Browning (e.g., pp. 209–216) offers a multilayered portrayal of the battalion during their first and subsequent mass killings. The largest group of policemen ended up doing whatever they were asked to, avoiding the risk of confronting authority or appearing to be cowards, yet not volunteering to kill. Increasingly numbed by the violence, they did not think that what they were doing was immoral, because it was sanctioned by authority. In fact, most tried not to think at all. A second group of "eager" killers who celebrated their murderous deeds increased in numbers over time. The smallest group were the nonshooters, who, with the exception of one lieutenant, neither protested against the regime nor reproached their comrades.

2. Some psychologists do invoke a "dual-process model" that postulates an "intuitive system" and a "reasoning system" to account for the difference between moral intuition and reasoning. In my opinion, however, this amounts to a redescription of the phenomenon rather than an explanation; contrary to what its name suggests, this model does not specify any process underlying intuition or reasoning but consists of a list of dichotomies (Gigerenzer & Regier, 1996).