7

Moral Intuitions

WALTER SINNOTT-ARMSTRONG, LIANE YOUNG,
AND FIERY CUSHMAN

Card players often find themselves in positions where they could cheat with little or no risk of being caught, but they don't cheat, because cheating strikes them as immoral. Similarly, needy customers are often able to steal an item that they want and cannot afford to buy, but they choose to go home empty-handed, because they see stealing as immoral. Many everyday decisions like these are based on moral intuitions.

So are moral theories. It is hard to imagine any way to develop a moral theory without relying on moral intuitions at all. How could you choose among consequentialism, Kantianism, contractarianism, and virtue theories without appealing to moral intuitions at some point in some way? Most contemporary moral theorists use something like the method of reflective equilibrium, which in effect systematizes moral intuitions, so they must admit that they rely on moral intuitions.

Because moral intuitions are crucial to everyday life and moral theory, they have attracted a great deal of attention from both philosophers and psychologists (as well as neuroscientists, more recently). Still, there is little agreement or conversation between philosophers and psychologists about moral intuitions. When they do discuss moral intuitions, it is not clear that they are talking about the same topic, since they often disagree on what counts as a moral intuition.

When we refer to *moral intuitions*, we shall mean strong, stable, immediate moral beliefs. These moral beliefs are *strong* insofar as they are held with confidence and resist counter-evidence (although strong enough counter-evidence can sometimes overturn them). They are *stable* in that they are not just temporary whims but last a long time (although there will be times when a person who has a moral intuition does not focus attention on it). They are *immediate* because they do not arise from any process that goes through

intermediate steps of conscious reasoning (although the believer is conscious of the resulting moral belief).

Such moral intuitions can be held about specific cases (e.g. a person, A, morally ought to keep this promise to this person, B), about general types of cases (e.g. whenever anyone promises to do anything, she or he morally ought to do it unless there is an adequate reason not to do it), or about very abstract principles (e.g. if A ought to do X, and A cannot do X unless A does Y, then A ought to do Y). We shall focus on moral intuitions about specific cases, because so little empirical research has been done on general or abstract moral intuitions.

Philosophers tend to ask *normative* questions about such intuitions: are they justified? When? How? Can they give us moral knowledge? Of what kinds? And so on. In contrast, psychologists tend to ask *descriptive* questions: how do moral intuitions arise? To what extent does culture influence moral intuitions? Are moral intuitions subject to framing effects? How are they related to emotions? And so on.

Until the last decade of the twentieth century, philosophers and psychologists usually engaged in their enterprises separately. This was unfortunate, because it is hard to see how to determine whether certain moral intuitions are justified without any understanding of the processes that produce those intuitions. We are not claiming that psychological findings alone entail philosophical or moral conclusions. That would move us too quickly from "is" to "ought." Our point is different: moral intuitions are unreliable to the extent that morally irrelevant factors affect moral intuitions. When they are distorted by irrelevant factors, moral intuitions can be likened to mirages or seeing pink elephants on LSD. Only when beliefs arise in more reputable ways do they have a fighting chance of being justified. Hence we need to know about the processes that produce moral intuitions before we can determine whether moral intuitions are justified. That is what interests us in asking how moral intuitions work.

There are several ways to answer this question. One approach is neuroscience (Greene et al., 2001, 2004). Another uses a linguistic analogy (Hauser et al., 2008). Those methods are illuminating, and compatible with what we say here, but they are discussed elsewhere in this volume. (See Chapters 2 and 8 in this volume.). Here we want to discuss a distinct, though complementary, research program. This approach is taken by psychologists who study heuristics and claim that moral intuitions are, or are shaped and driven by, heuristics.

A few examples of non-moral heuristics will set the stage. Then, after identifying a general pattern, we can return to ask whether moral intuitions fit that pattern.

1. Non-Moral Heuristics

How many seven-letter words whose sixth letter is ''n'' (____n_) occur in the first ten pages of Tolstoy's novel, *War and Peace*? Now, how many seven-letter words ending in ''ing'' (____ing) occur in the first ten pages of *War and Peace*? The average answer to the first question is several times lower than the average answer to the second question. However, the correct answer to the first question cannot possibly be lower than the correct answer to the second question, because every seven-letter word ending in ''ing'' is a seven-letter word whose sixth letter is ''n''. Many subjects make this mistake even when they are asked both questions in a single sitting with no time pressure. Why? The best explanation seems to be that their guesses are based on how easy it is for them to come up with examples. They find it difficult to produce examples of seven-letter words whose sixth letter is ''n'' when they are not cued to think of the ending ''ing''. In contrast, when asked about seven-letter words ending in ''ing'', they easily think up lots of examples. The more easily they think up examples, the more instances of the word-type they predict in the ten pages. This method is called *the availability heuristic* (Kahneman et al., 1982, chs. 1, 11–14). When subjects use it, they base their beliefs about a relatively inaccessible attribute (the number of words of a given type in a specified passage) on a more accessible attribute (how easy it is to think up examples of such words).

A second classic heuristic is *representativeness*. Kahneman et al. (1982: ch. 4) gave subjects this description of a graduate student:

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.

Subjects were given a list of nine fields of graduate study. Subjects in one group were then asked to rank those fields by the degree to which Tom ''resembles a typical graduate student'' in each field (and, hence, is representative of that field). Subjects in another group were asked to rank the fields by the likelihood that Tom is in each field. Both groups of subjects were also asked to estimate the percentage of graduate students in each of the nine fields. These estimates varied from 3% to 20%, and Tom's description fit the stereotype of the smaller fields, such as library science, but not larger fields, such as English.

These percentage estimates should have big effects on subjects' probability rankings, because any given graduate student is less likely to be in a field that is smaller. Nonetheless, subjects' percentage estimates had almost no effect on their probability rankings. Instead, the answers to the questions about representativeness and probability were almost perfectly correlated (0.97). This suggests that these subjects neglected the baseline percentage and based their probability estimates almost totally on their judgments of representativeness. As before, they substituted a relatively accessible attribute (representativeness) for a relatively inaccessible attribute (probability).[1]

FN:1

A third example is the *recognition heuristic*, studied by Gigerenzer et al. (1999: chs. 2–3). When asked which US city (San Diego or San Antonio) or which German city (Berlin or Munich) is larger, people tend to guess cities they recognize. This heuristic makes sense on the reasonable assumption that people hear more about bigger cities. Still, this heuristic can also be misleading. Gigerenzer's group found that subjects followed the recognition heuristic regularly (median 100%, mean 92%), even after they received information that should lead them to stop following this decision rule, such as information about which cities have professional soccer teams (1999: 50–52). Again, these subjects seem to base their beliefs about a relatively inaccessible attribute (population) on an accessible attribute (recognition) rather than on other available information that is known to be relevant.

1.1. *Battle of the Titans*

We included examples from both Kahneman and Gigerenzer because their research programs are often seen as opposed. Gigerenzer and his colleagues emphasize that simple heuristics can make us smart, whereas Kahneman, Tversky, and their colleagues study how heuristics and biases can lead to mistakes. However, this difference is largely a matter of emphasis (Samuels, Stich, & Bishop, 2002). Both sides agree that our heuristics lead to accurate enough judgments in most cases within typical environments. Otherwise, it would be hard to understand why we evolved to use those heuristics. Both sides also agree that heuristics can lead to important mistakes in unusual environments, and they agree that which heuristics lead to mistakes in which environments is a matter for empirical research.

Kahneman and Gigerenzer might still seem to disagree about rationality. Gigerenzer argues that it is rational to employ heuristics because heuristics

[1] A better-known example of representativeness is Linda the feminist bank teller (Kahneman et al., 1982: ch. 6). This example is controversial, because some critics claim that, in a list with "feminist bank teller," "bank teller" alone might be interpreted as "non-feminist bank teller." We will avoid the controversy by discussing the case of Tom, which is not subject to this objection.

provide the best method available in practice. In contrast, Kahneman suggests that people who use heuristics exhibit a kind of irrationality insofar as their responses violate rules of logic, mathematics, and probability theory. Again, however, we doubt that this disagreement is deep, since the apparently conflicting sides use different notions of rationality, and neither can legitimately claim that their notion of rationality is the only defensible one. If a heuristic is the best available method for forming true beliefs, but sometimes it leads people to violate the rules of logic, math, or probability, then it is rational in Gigerenzer's practical sense to use the heuristic even though this use sometimes leads to irrationality in Kahneman's formal sense. They can both be correct.

Gigerenzer and his followers also complain that Kahneman's heuristics are not specified adequately. They want to know which cues trigger which particular heuristic, which computational steps run from input to output, and how each heuristic evolved. We agree that these details need to be spelled out. We apologize in advance for our omission of such details here in this initial foray into a heuristic model of moral intuition. Still, we hope that the general model will survive after such details are specified. Admittedly, that remains to be shown. Much work remains to be done. All we can do now is try to make the general picture attractive and show its promise.

1.2. *Heuristics as Unconscious Attribute Substitutions*

What is common to the above examples that makes them all heuristics? On one common account (Sunstein, 2005), heuristics include any mental short-cuts or rules of thumb that generally work well in common circumstances but also lead to systematic errors in unusual situations. This definition includes explicit rules of thumb, such as "Invest only in blue-chip stocks" and "Believe what scientists rather than priests tell you about the natural world." Unfortunately, this broad definition includes so many diverse methods that it is hard to say anything very useful about the class as a whole.

A narrower definition captures the features of the above heuristics that make them our model for moral intuitions. On this narrow account, which we shall adopt here, all heuristics work by means of unconscious attribute substitution (Kahneman & Frederick, 2005).[2] A person wants to determine whether an object, X, has a target attribute, T. This target attribute is difficult to detect

FN:2

[2] This pattern is not shared by all methods that are called "heuristics" by psychologists. Some exceptions, such as anchoring, still resemble paradigm heuristics in specifiable ways (Kahneman & Frederick, 2005: 272). Other so-called heuristics enable action or decision without belief. Gigerenzer discusses the gaze heuristic that baseball outfielders follow to catch a fly ball: fixate your gaze on the ball, start running, and adjust your speed so that the angle of gaze remains constant (2008: 7). Outfielders who use this heuristic rarely form any belief about the heuristic, about why they use it, about what the

directly, often due to the believer's lack of information or time. Hence, instead of directly investigating whether the object has the target attribute, the believer uses information about a different attribute, the heuristic attribute, H, which is easier to detect.[3] The believer usually does not consciously notice that he is answering a different question: "Does object, X, have heuristic attribute, H?" instead of "Does object, X, have target attribute, T?" The believer simply forms the belief that the object has the target attribute, T, if he detects the heuristic attribute, H.

In the above case of availability, the target attribute is the rate of occurrence of certain words, and the heuristic attribute is how easy it is for this person to think up examples of such words. In the above case of representativeness, the target attribute is the probability that Tom is studying a certain field, and the heuristic attribute is how representative Tom's personal attributes are of each field. In the above case of recognition, the target attribute is a city's population, and the heuristic attribute is ease of recognizing the city.

In some of these cases, what makes the heuristic attribute more accessible is that it is an attribute of the person forming the belief rather than an attribute of the object. How easy it is for someone to think up certain words or to recognize the name of a city is a property of that person. In contrast, the target attribute is not an attribute of the person. In our examples, it is an attribute of words, cities, and Tom. Thus the heuristic attribute need not be an attribute of the same thing as the target attribute.

Nonetheless, these heuristic attributes are contingently and indirectly related to their target attributes. In some cases, the heuristic attribute is even a part of the target attribute. For example, in "one-reason" decision-making (Gigerenzer et al. 1999: chs. 4–8), we replace the target attribute of being supported by the best reasons overall with the heuristic attribute of being supported by a single reason. When people buy cars in this way, they might focus on efficiency alone instead of trying to think about all of the pros and cons of all available cars at once. Why do people focus on a single reason? It is too difficult to consider all of the many and varied considerations that are relevant, and too much information can be confusing or distracting, so people are often more accurate when they consider only one reason. Then they base their decision on the simpler heuristic attribute rather than on the more complex target attribute.

angle is, and so on. This heuristic helps them catch balls, not form beliefs. In contrast, we shall focus on heuristics that help people form beliefs or judgments.

[3] To say that subjects substitute one attribute for another is not to say that they would treat them identically in all respects if explicitly probed. The point is just that subjects answer a question about the target attribute by answering a question about the heuristic attribute.

Heuristics come in many forms, but all of the heuristics discussed here involve unconscious attribute substitution. This account applies to a wide variety of heuristics from Kahneman, Gigerenzer, and others (such as Chaiken, 1980, who discusses the "I agree with people I like" heuristic, and Laland, 2001, who adds the "do what the majority does" heuristic). Unconscious attribute substitution is what makes them all heuristics in the narrow sense that will concern us here.

1.3. *Heuristics are not just Old-Fashioned Inferences from Evidence*

Such unconscious attribute substitution might seem to involve some form of inference. The believer moves from a belief that the object has the heuristic attribute to a belief that the object has the target attribute. This process starts with a belief and adds a new belief, so it might seem to be an inference. The heuristic attribute might even seem to be evidence for the target attribute.

We have no objection to these labels. Still, they should not hide the important differences between heuristics and what is commonly seen as evidence.

First, heuristics normally operate unconsciously. Some people might consciously appeal to availability, representativeness, or recognition in order to answer a question. However, most characteristic uses of heuristics are unconscious. This is shown in several ways. Subjects in the reported experiments usually do not mention the heuristic attribute when asked to explain how they arrived at their answers. In contrast, imagine the description said that Tom W. reads books about library science. Then subjects would, presumably, cite his reading habits as evidence when they are asked why they believe that Tom W. is in library science. If they did not report his reading habits, this omission would suggest that they did not use his reading habits as evidence for their beliefs, assuming they had no special reason to avoid mentioning this evidence. Similarly, when subjects do not mention the heuristic attribute of representativeness, that omission suggests that subjects do not use that attribute as evidence in the same way as they would use his reading habits as evidence.

Furthermore, unconscious processes are disturbed less by concurrent cognitive loads on working memory (such as distracting irrelevant tasks) than conscious processes are. Thus, if heuristics are unconscious but can be monitored and corrected by conscious processes, as dual-process models suggest, then subjects with greater cognitive loads will deviate less from the heuristic even when it is obviously mistaken. That's exactly what is found in experiments (Kahneman & Frederick, 2005, 268, 273, 285). In contrast, concurrent cognitive loads would seem to increase deviations from inferences based on evidence. All of that suggests that heuristics do not operate consciously in the way that normal inferences and evidence do.

Second, partly because heuristics are unconscious, they not easily corrected when they go astray. Sure enough, researchers find that even experts on probability make the mistakes predicted by the various heuristics. When the experimental design makes the mistakes obvious enough, and there is no concurrent cognitive load, then experts do make fewer salient mistakes (Kahneman & Frederick, 2005: 273, 278–9, 287). Still, the persistence of heuristics is remarkable. In contrast, experts find it easier to correct misleading evidence, partly because the evidence exerts its effects through conscious reasoning.

Third, attribute substitution plays a role that normal evidence does not, insofar as attribute substitution silences or excludes or distracts from opposing evidence. If Tom W. is reported to wear a hat emblazoned with ''Milton Matters'', subjects would presumably weigh this evidence against his personal attributes and reduce their estimates of the probability that Tom W. is in library science instead of English. In contrast, when representativeness is substituted for probability in Kahneman's case of Tom, for example, representativeness is not weighed against the other evidence coming from subjects' percentage estimates. Instead, the baseline percentages are overlooked, and the judgment is based almost completely on the heuristic attribute of representativeness. Gigerenzer got similar results for his recognition heuristic, as discussed above. This silencing of other considerations might be appropriate: sometimes too much information can increase rates of errors. In any case, their role as silencers shows that heuristics have a force that normal evidence (such as Tom's hat) lacks.

Some philosophers allow unconscious evidence and exclusionary reasons, so they still might insist on calling the heuristic attribute evidence. We do not want or need to argue about whether heuristic attributes really are or are not evidence. If heuristic attributes are evidence, they are a special kind of evidence that operates in a special way with a special force. That is what matters here.

1.4. *Do we use Non-Moral Heuristics?*

Whether or not we classify heuristics as inferences or as evidence, the important questions are when, in which ways, and to what extent heuristics guide our beliefs. How can we tell?

The most direct evidence for attribute substitution comes from correlations between answers to questions about the target attribute and about the heuristic attribute. In the case of the representativeness heuristic, as we said, subjects' answers to questions about representativeness and probability were almost perfectly correlated (0.97). The near perfection of this correlation strongly suggests

that these subjects substituted the heuristic attribute of representativeness for the target attribute of probability in answering questions about the target attribute.

If subjects were answering a question about probability and using representativeness only as standard evidence for probability, then their answers would reflect other factors, including counter-evidence against their probability estimates. This counter-evidence might be overridden, but it would not be silenced in the sense of removing all of its force. In particular, their answers would be expected to vary with baseline percentages, since baseline percentages affect probabilities. But then the correlation between judgments of representativeness and of probability would not be as high as they are, because probability would vary with baseline percentage but representativeness would not. The near perfection of the observed correlations thus suggests that subjects do not simply treat heuristic attributes as one bit of evidence among others.

Instead, these high correlations suggest that subjects do not even distinguish the questions of probability and representativeness. If subjects were answering questions about representativeness, then they would be expected to ignore baseline percentages, because representativeness is not affected by how many students are in a field. The observed high correlation would, thus, be expected if subjects answered a question about representativeness when asked about probability. By substituting one question for another, they silence all of the factors or counter-evidence that affect the target attribute but not the heuristic attribute. The attribute substitution hypothesis thus explains the observed high correlations.

The attribute substitution hypothesis also explains some framing effects. When several heuristic attributes might be substituted for a given target attribute, when those different heuristics would lead to different beliefs about the target attribute, and when contextual framing affects which of these heuristic attributes gets substituted for the target attribute, then contextual framing will affect subjects' beliefs about the target attribute (Kahneman & Frederick, 2005: 269) Subtle wording and order differences can trigger different heuristics and thereby lead to different judgments. This mysterious phenomenon of framing thus receives a natural explanation from the hypothesis that these beliefs are driven by heuristics.

Much more could be said in favor of an attribute substitution account of non-moral heuristics, but the general pattern should be clear, and there is abundant evidence that this pattern is common outside morality. The next question is whether the use of heuristics in the form of attribute substitution also occurs in the domain of morality.

2. Moral Heuristics

Moral intuitions fit the pattern of heuristics, in our "narrow" sense, if they involve (a) a target attribute that is relatively inaccessible, (b) a heuristic attribute that is more easily accessible, and (c) an unconscious substitution of the target attribute for the heuristic attribute. We shall discuss these elements in turn.

2.1. *Is the Target Attribute Inaccessible?*

The target attribute in a moral judgment is simply the attribute that the person who makes the judgment ascribes to the act, correctly or not.[4] When someone judges that an act is morally wrong, the target attribute for that judgment is moral wrongness. When judging that someone is morally virtuous, the target attribute for that judgment is moral virtue.[5] Similarly for judgments of moral goodness, rights, and so on. Each of these target attributes is relatively inaccessible in its own way, but in the interest of streamlining our discussion, we shall focus on moral wrongness as a target attribute.

Many people seem to think that they have easy and direct access to moral wrongness. They claim to see that acts are wrong. However, that's also what people think when they feel confident that Tom is more likely to be in library science than in English because he is more representative of the former type of graduate student. A target attribute (such as probability) can be difficult to access directly, even if it seems easily accessible to people who confuse it with a heuristic attribute that is easily accessible. Analogously, any impression that we can easily access moral wrongness by means of direct moral intuition might be explicable as a confusion of the target attribute with its heuristic. Thus apparent accessibility cannot show that moral wrongness has real accessibility.

To determine whether moral wrongness really is accessible, we need to ask what accessibility is and what moral wrongness is. With regard to heuristics, the relevant notion of accessibility is accessibility in practice, with realistic constraints on time and information. After all, the point of heuristics is to be fast and frugal enough to be useful in real life.

[4] In this general sense, even expressivists who are quasi-realists can talk about target attributes.

[5] Consider moral virtue. Evidence for situationism is sometimes said to make improbable the existence of robust character traits, such as virtues (Doris, 1998, 2002; Harman, 1999; Merritt, 2001; Merritt et al., Chapter 11, this volume). We need not go so far in order to recognize that it is no simple matter to tell who has a moral virtue or vice. Such judgments depend not only on how agents would act in circumstances that they have never faced but also on their internal motivations, which are hard to tell from the outside (and often from the inside).

To determine whether moral wrongness is accessible in this way, we need to ask what moral wrongness is. Luckily, we do not need to commit ourselves to any particular account of moral wrongness, because the plausible candidates suggest that moral wrongness is not accessible in the relevant way.[6] Consider the consequentialist view that whether an act is morally wrong depends only on whether some alternative has better consequences overall. It is notoriously hard to tell which act maximizes pleasure and pain, much less the good. This requires knowing far into the future, and nobody has the information or capacity required to calculate the total or average. Thus, if the attribute of moral wrongness is the attribute of failing to maximize the good, then this target attribute is definitely inaccessible.[7] It does not help much to make moral wrongness depend only on expected value, since it is also often hard to tell whether a real agent reasonably believes that an act will maximize the good.

Kantianism might seem to make moral wrongness easier to access in real life, but it doesn't. The ongoing debates about Kant's first formula of the categorical imperative show that it is hard to tell what the maxim of an act is, what it means for an act to be universalizable or not, and whether a given maxim is universalizable. It is also hard to say exactly what it is to treat someone as a means only, as prohibited by Kant's second formula, and whether the act would violate the rules of a kingdom of ends, as prohibited by Kant's third formula. The general point is that such theoretically delicate notions are too complex and unwieldy for moral decisions by common folk in normal situations.

The same goes for contractarianism, as in Rawls (1971) and Scanlon (1999). If to judge an act wrong is to judge that it violates rules that all rational impartial people would accept or, instead, rules that no reasonable person could reject, then it will be very hard for any ordinary person in real life, likely to be both

[6] Although we focus on substantive moral theories in the text, meta-ethical theories support the same conclusion. In speaking of moral wrongness as a target property, we presuppose realism or at least quasi-realism. Cornell moral realists (e.g. Brink, 1989) argue that moral wrongness is a natural kind, just as water is H_2O. The chemical compositions of liquids are not accessible directly, so we need to use their phenomenal properties to identify liquids in everyday life. If moral wrongness is analogous, we shall not have direct access to the natural property that is moral wrongness. Cornell realists usually say that moral wrongness is a homeostatic cluster and is discovered by a complex process of reflective equilibrium, both of which suggest inaccessibility in everyday life. Inaccessibility is also suggested by Canberra moral realists, who analyze moral wrongness as the property that best explains a large set of shared platitudes about moral wrongness (e.g. Jackson, 1998: chs. 5–6). It would clearly be very difficult for common people to think so systematically about moral wrongness when making moral judgments in everyday life. Admittedly, moral realists could claim that moral wrongness is a simple non-natural property that is accessible directly (e.g. Ross, 1930). But how is it accessed? And why adopt such a queer metaphysics and epistemology?

[7] This inaccessibility in practice is not a problem for consequentialists if they claim only a theoretical standard or criterion of rightness instead of a practical guide or decision procedure (see Bales, 1971).

irrational and partial, to determine directly whether any act is morally wrong. Experts might be able to apply such abstract theories after long reflection, but that cannot be what is going on in everyday life.

Social moral relativism might seem to aid access to moral wrongness, if it makes moral wrongness depend on the actual conventions of society (Harman, 1996). However, it is not easy or quick for someone judging a particular act to determine which conventions are essential to the moral wrongness of this act and also whether those conventions are in place in this society, because which conventions exist in a society depends on patterns of action and motivation for many people other than the person making the moral judgment.

There are many other possibilities, but, in the end, the only theories that make moral wrongness easily accessible are those that identify moral wrongness with the judger's own emotional reactions. However, such subjectivism is totally implausible, as was shown long ago, because of its inability to account for interpersonal disagreements and other common features of morality. Thus no plausible theory will make moral wrongness accessible in practice without heuristics.

2.2. *Which Moral Heuristic?*

Inaccessibility creates the need for a heuristic attribute. If moral wrongness were easily accessible, we would not need to substitute any heuristic attribute in order to judge whether an act is morally wrong. Because moral wrongness is so hard to access directly, we need to substitute a more easily accessible heuristic attribute in order to be able to judge whether an act is morally wrong.

Which heuristic attribute? A heuristic attribute must be easily and quickly accessible (like availability, representativeness, or recognition). Heuristics are supposed to be fast and frugal, after all, and reliable, at least in common situations. A heuristic attribute must, therefore, be related somehow to the target attribute. Otherwise, it would be a mystery why we evolved to substitute that heuristic. Still, there are lots of relevant accessible attributes that people might substitute for the inaccessible attribute of moral wrongness.

One possibility, as Gigerenzer (2008: 9) puts it, is that "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." Heuristics that guide non-moral beliefs, decisions, and actions clearly also affect moral beliefs, decisions, and actions. Gigerenzer mentions Laland's (2001) do-what-the-majority-do heuristic: if you see the majority of peers behave in a certain way, do the same.

We could add Chaiken's (1980) I-agree-with-people-I-like heuristic. These heuristics affect moral and non-moral beliefs alike. However, it seems unlikely that all moral beliefs and actions can be explained completely in terms of general heuristics that apply outside as well as inside morality. Morality sometimes leads us to criticize the majority as well as people we like, and moral judgments are not entailed simply by factual beliefs that ordinary heuristics enable, so there must be additional heuristics behind at least some moral intuitions.

Another suggestion is that the heuristic attributes behind common moral judgments are the attributes mentioned in common moral rules and principles, such as don't kill, disable, cause suffering, lie, cheat, steal, or break promises, at least without an adequate justification. People might seem to use these categories to reach judgments about moral wrongness. Instead of directly asking whether an act is morally wrong, they might seem to classify the act as killing, say, and then infer that it is morally wrong because it is killing.

However, recent experimental evidence suggests that people do not form moral judgments by applying a rule about killing or by checking whether the act has the attribute of being a killing (Sinnott-Armstrong et al., 2008; described in Chapter 6 in this volume). Other studies suggest that we often cannot apply the notion of causation, which is central to several of the common moral rules, without presupposing some prior moral judgment (Alicke, 1992). If we cannot apply common moral rules without prior moral judgments, then an account of moral intuitions needs to explain those prior moral judgments and not just the way that they get used in applying common moral rules.

More generally, this approach is too indeterminate. You can view any moral principle as a heuristic. Just-war theory, for example, includes a principle that forbids preventive war as opposed to pre-emptive war. On this view, the property of being preventive is what makes a war immoral. In contrast, this principle can be reconceived as a heuristic where the target attribute is moral wrongness and the heuristic attribute is being preventive war. On this heuristic view, the property of being preventive does not itself make war wrong but instead is substituted for the separate target property of wrongness. So, which is it: principle or heuristic? There is no solid basis for either claim without substantive moral assumptions about what makes wrong acts wrong. This problem generalizes to other principles. Sunstein (2005) writes, ''what a utilitarian sees as a moral heuristic (never lie!) might be regarded as a freestanding moral principle by a deontologist.'' Deontologists can even see the basic principle of utility as a heuristic that usually works well but leads us astray in exceptional cases. There seems to be no way to tell which is a moral principle and which is a heuristic without making assumptions about what is morally wrong and why. Scientists need a more neutral way to study heuristics.

Finally, these principles might be heuristics in the broad sense mentioned above, but they are normally conscious, so they are crucially different from the narrow heuristics that concern us here. Subjects who use the availability heuristic are rarely conscious of applying any rule like, ''If you can think of lots of examples, then guess that there will be lots more.'' That makes it misleading to see the features in conscious moral rules as heuristic attributes.

A third set of heuristic attributes is proposed by Sunstein (2005):

> *Cold-heart Heuristic*: ''Those who know they will cause a death and do so anyway are regarded as cold-hearted monsters'' (This is supposed to explain widespread opposition to cost–benefit analysis.)
>
> *Fee Heuristic*: ''People should not be permitted to engage in wrongdoing for a fee.'' (This is supposed to explain opposition to emissions trading.)
>
> *Betrayal Heuristic*: ''Punish, and do not reward, betrayals of trust.'' (This is supposed to explain opposition to safer airbags that cause some deaths.)
>
> *Nature Heuristic*: ''Don't tamper with nature'' (a.k.a. ''Don't play God'') (This is supposed to explain some opposition to genetic engineering and cloning.)
>
> *Action Heuristic*: ''Don't do harmful acts.'' (This is supposed to explain why people see doing harm as worse than allowing harm, as in active *vs*. passive euthanasia and in vaccination policies.)

These might all count as heuristics under Sunstein's broad definition. However, just as with the common moral rules mentioned above, there is no morally neutral way to tell whether these are heuristics or new deontological rules. Also, Sunstein's heuristics cannot be moral heuristics in our narrow sense, because they typically operate consciously rather than unconsciously.

Some other moral principles do seem to operate unconsciously. Subjects often make moral judgments in a pattern prescribed by the doctrine of double effect without being able to cite that principle as a justification and, presumably, without being conscious of using that principle (Hauser et al., 2008; Cushman et al., 2006; Mikhail, 2002). The attribute that makes an act violate the doctrine of double effect is that its agent intends harm as a means. This attribute might, therefore, operate much like heuristic attributes. That's why moral judgments that fit that pattern and are based on that principle are classified as moral intuitions, in contrast to moral judgments that are inferred from conscious rules.

Other factors also unconsciously affect our moral intuitions. Most people, for example, are more likely to judge harmful acts as morally wrong when their agents touch the victim and less likely to judge acts morally wrong when

the victim is at some physical distance. When asked whether contact and physical distance are morally important, they often deny that it is important or at least admit that they cannot say why it is important (Cushman et al., 2006). This suggests that contact and physical distance might be operating as heuristic attributes for the target attribute of moral wrongness.

However, the attribute of intention as a means is not so easy to access. And recent research (Greene et al., 2009) suggests that the contact principle needs to be reformulated as a less accessible principle of personal force that applies when one transfers energy to another directly by the force of one's muscles. This inaccessibility makes these attributes bad candidates for heuristic attributes, which are supposed to be easily accessible.

One promising approach tries to account for a variety of moral intuitions, as well as the grab-bag of moral heuristics proposed above, as instances of a general *affect heuristic* (Kahneman & Frederick, 2005: 271b, 283; cf. Slovic et al., 2002). We shall sketch this approach here and discuss evidence for it in a subsequent section (2.3). Unlike the aforementioned moral heuristics, which caution against specific acts or act-types, the affect heuristic is content-free. All the affect heuristic says is, roughly: if thinking about the act (whatever the act might be) makes you feel bad in a certain way, then judge that it is morally wrong. The point, of course, is not that everyone consciously formulates this conditional. The claim is only that people unconsciously substitute how the act makes them feel for the target attribute of moral wrongness and then judge the act morally wrong on the basis of how it makes them feel.[8]

FN:8

The relevant bad feelings are diverse: if you consider doing the act yourself, you might feel compunction in advance or anticipate guilt and/or shame afterwards. If you imagine someone else doing the act to you, then you might feel anger or indignation. If you imagine someone else doing the act to a third party, then you might feel outrage at the act or the agent. (The outrage heuristic of Sunstein, 2005 is, thus, a part of the larger affect heuristic.) And different kinds of negative affect might accompany judgments in different areas of morality: anger in moral judgments about harm, contempt in the moral judgments about hierarchy, and disgust in moral judgments about impurity. (Cf. Haidt & Joseph, 2004 on the CAD hypothesis and Kass, 1997 on the "wisdom of repugnance".) In all these cases, some kind of negative affect accompanies some kind of moral judgment. The affect operates as a heuristic

[8] The affect need not be felt every time the judgment is made. Instead, it could be made in some cases and then generalized into a rule, or it could be felt in paradigm cases and then lead us to judge other cases wrong because they resemble the paradigm cases.

attribute if (but only if) people reach the moral judgment by unconsciously substituting the affect for the target attribute of moral wrongness.

This affect heuristic might underlie the other moral heuristics. If people feel worse when they imagine causing harm intentionally than when they imagine causing harm unintentionally (cf. Schaich Borg et al., 2006), then the hypothesis that people follow an affect heuristic predicts that moral judgments will display the pattern prescribed by the doctrine of double effect. Similarly, if we feel worse when we imagine causing harm to someone we contact (or exert personal force on), then the hypothesis that we follow an affect heuristic predicts that our moral judgments will follow the pattern predicted by the contact (or personal force) heuristic. Similarly for other heuristics. And people do seem to feel bad when they imagine acts that violate such heuristics. Thus the affect heuristic might explain what is common to all (or many) of the other postulated moral heuristics.

Much work needs to be done in order to determine which cues trigger which emotions and, thereby, which moral judgments. As long as the process works by means of emotion or affect, however, it will be illuminating to see moral intuitions as based on an affect heuristic.

2.3. *Moral Heuristics and Biological Adaptation*

The preceding accounts of moral heuristics as attribute substitutions might seem to presuppose that moral wrongness is a real attribute or property. Hardened skeptics, however, might deny that moral wrongness is any property of events to begin with. They might complain that moral intuitions cannot operate as heuristics, because there is no objective moral truth for them to approximate. They might add that there is no moral reality beyond moral intuitions as psychological objects. In this section we suggest that moral intuitions can be fruitfully understood as heuristics even on such a skeptical view. In doing so, we rely on the assumption that moral intuitions have an important functional role as biological adaptations.

Many of the above examples of non-moral heuristics apply in cases where an individual is attempting to establish some sort of factual representation, such as the number of words of a particular type in a text, the probability of a person possessing a set of attributes, or the population of a city. In each of these cases, the output of the heuristic is a mental representation of some feature of the world.

One purpose of representing the world, of course, is to guide action. Thus we might substitute attributes (''he has a shaved head, spiked leather chains and lots of tattoos'') for a target property (''he is dangerous'') in order to

produce the most appropriate course of action (''move away from him''). Importantly, we can sometimes construct heuristics in such a way that they move directly from the substitute attribute to the appropriate course of action without ever representing the target property. We might, for instance, tell our children ''Avoid men with tattoos'' without bothering to explain that we are substituting certain visual correlates for the underlying target property of dangerousness. For the sake of clarity, let's call heuristics of this second type ''motivational heuristics'' and contrast them with ''representational heuristics.'' Representational heuristics output representations of some fact about the world, while motivational heuristics output the appropriate course of action.

Natural selection makes abundant use of motivational heuristics. Bad-tasting food directly motivates an avoidance response without any direct representation of the underlying probability of toxic or pathogenic qualities. Good-looking people directly motivate an approach response without any direct representation of the underlying probability of fecundity or fitness. Moreover, there is a sense in which ''bad-tasting''or ''good looking'' do not refer to a property of foods or people in any objective sense at all. These are psychologically constructed attributions that operate as heuristic approximations of some very real underlying attributes, but which motivate particular behaviors directly, without generating representations of the underlying attributes.

Moral intuitions could be understood in much the same way. Moral judgment motivates a host of behaviors: do this, don't do that; punish him, reward her; shun her, affiliate with him. Prosocial behaviors are adaptive because they help individuals to reap the payoffs of cooperative interactions, to avoid sanctions, and to enforce prosociality among social partners. People often perform behaviors motivated by moral intuitions without directly representing the underlying reason why the behavior is adaptively favorable.

One of the most compelling cases for adaptive moral heuristics is incest avoidance. Taboos against sexual intercourse between first-degree relatives come as close to cross-cultural universality as any social psychological phenomenon (Wilson, 1998). Indeed, reproduction between first-degree relatives is uncommon among most sexually reproducing species. There is an obvious adaptive explanation for this behavior: reproduction between first-degree relatives tends to produce less biologically fit offspring, especially over successive generations. Our moral intuition that incest is wrong is not a representational heuristic for estimating the fitness consequences of reproduction, but it can be fruitfully understood as a motivational heuristic for behavior that tends to have fitness-enhancing consequences.

We're now in a position to re-evaluate whether skepticism about moral facts threatens our proposal to understand moral intuitions as heuristics. If we regard

moral intuitions as representational heuristics—in particular, if we regard moral intuitions as heuristic devices for estimating the moral truth in a computationally efficient manner—then the moral skeptic will have little use for our proposal. On the other hand, if we regard moral intuitions as a motivational heuristic for producing adaptive behaviors, skepticism about moral facts is not threatening at all. We may regard moral intuitions as subjective psychological states that exist because they motivate fitness-enhancing behaviors in a computationally efficient manner.

2.4. *Do we use Moral Heuristics?*

Of course, it is easy to postulate and hard to prove. We need evidence before we can conclude that we really do follow the affect heuristic or any other heuristic when we form moral judgments. Since heuristics are unconscious attribute substitutions, the evidence comes in two stages: we need evidence that the process is attribute substitution. Then we need evidence that the process is unconscious. We also need some reason to believe that the heuristic attribute is related in a significant way to the target attribute in order to understand why that heuristic arose.

What is the evidence for attribute substitution? As with non-moral heuristics, the most direct evidence of attribute substitution would be strong correlations between answers to questions about the target and heuristic attributes. Which heuristics are at work will be revealed by which correlations hold.

Consider the affect heuristic. When subjects presented with various harmful acts were asked (a) how ''outrageous'' the act was and also (b) how much the agent should be punished, the correlation between their two answers was a whopping 0.98 (Kahneman et al., 1998). The term ''outrageous'' might not signal emotion, but a later study by Carlsmith et al. (2002) asked subjects how ''morally outraged'' they were by the act and then how severe the crime was and how much the agent should be punished. The correlations between outrage and severity and between outrage and sentence, respectively, were 0.73 and 0.64 in their study 2 and 0.52 and 0.72 in their study 3. These correlations, though significant and high, are not as high as in Kahneman et al., but that reduction might be because the crimes in Carlsmith et al. were milder. In any case, these high correlations are striking because people often refer to consequences, such as deterrence, when asked abstractly why societies should punish, but such consequences seem to have little effect on their judgments of how much to punish in concrete cases. This conflict suggests that people might have both an unconscious emotional system that drives their concrete moral judgments and a conscious consequentialist system that drives their

abstract moral judgments (see Greene et al., 2001, 2004, 2008; Cushman et al., 2006 proposes a similar view). This dichotomy in moral judgments seems analogous to subjects' tendency to say that the baseline is crucial to probability when asked abstractly but then to overlook baseline percentages and focus solely on representativeness when asked to make a probability judgment in a concrete case.

A correlation between affect and moral judgment has also been found in a different area of morality. Haidt et al. (1993) presented their subjects with offensive actions, including eating one's pet dog who had died in an accident, masturbating in a chicken carcass and then eating it, and so on. Then they asked whether such acts are harmful, whether they are morally wrong, and whether it would bother them to witness the action. In the groups that tended to judge these acts morally wrong, the correlation between this moral judgment and predictions of bothersome affect was 0.70 (which was higher than the correlation with their answers to the harm question). This relation held across cultures (Brazil and Philadelphia). In contrast, moral judgments were more highly correlated with judgments of harm than with judgments of bothersome affect in those groups (especially higher socioeconomic Philadelphians) that tended to judge that these offensive acts are not morally wrong. On the assumption that people in these groups are better educated, this variation might reflect a greater tendency to control the unconscious emotional system with the less emotional conscious system, just as better-educated subjects tended to make fewer mistakes in the experiments with non-moral heuristics. Again, moral intuitions resemble non-moral heuristics in important respects.

Others have found many more correlations with affect. See Sunstein (2005), Kahneman and Frederick (2005: 271b, 283), and the neuroimaging studies of Greene et al. (2001, 2004) and Schaich Borg et al. (2006). Still, these are only correlations. All of these studies leave open the possibility that emotion is an after-effect of moral judgment. However, that possibility is undermined by a steadily increasing number of studies using diverse methods from both psychology and neuroscience.

One approach is to manipulate the extent of emotional engagement during moral judgment. For example, Haidt and colleagues have boosted emotional engagement by either hypnosis or priming. Wheatley and Haidt (2005) showed that when highly hypnotizable individuals were given a posthypnotic suggestion to experience disgust upon encountering an arbitrary word, they made harsher judgments of both morally relevant actions (e.g. eating one's dead pet dog, shoplifting) and morally irrelevant actions (e.g. choosing topics for academic discussion) specifically when these actions were described in vignettes including the disgust-inducing word.

Governed by the same logic, a second study (Schnall et al., 2008) probed subjects' responses to moral scenarios featuring morally relevant actions such as eating one's dead pet dog while priming subjects to feel disgusted. In one experiment, subjects filled out their questionnaires while seated at either a clean desk or a disgusting desk, stained and sticky and located near an overflowing waste bin containing used pizza boxes and dirty-looking tissues. Subjects who were rated as highly sensitive to their own bodily state were more likely to condemn the actions when seated at the disgusting desk than at the clean desk. Such studies suggest that emotion is not epiphenomenal when it comes to moral judgments but can instead causally affect moral judgments.

Greene and colleagues (2008) employed a more indirect manipulation of emotion via cognitive load, thought to interfere with "controlled cognitive processes", and therefore to result in a relative increase in the emotional contribution to moral judgment. As predicted, cognitive load slows down consequentialist responses but has no effect on nonconsequentialist responses, suggesting that affect plays a causal role in at least some moral judgments, specifically, nonconsequentialist responses.

In contrast to the previous studies, Valdesolo and DeSteno (2006) sought to *reduce* affect, specifically, negative affect, by presenting short comedic film clips to subjects before they produced moral judgments. Reducing negative affect was found to result in a greater proportion of consequentialist judgments, supporting the proposal that (negative) affect is not merely associated with but critically drives nonconsequentialist judgments.

Finally, studies of clinical populations have provided similar support for the affect heuristic or, at least, the necessary role of affect in moral judgment. Mendez and colleagues (2005) showed that patients with frontotemporal dementia and resulting blunted affect produce more consequentialist moral judgments as compared to healthy subjects. Koenigs, Young, and colleagues (2007) and Ciaramelli and colleages (2007) found a similar pattern of heavily consequentialist moral judgments among patients with severe emotional deficits due to ventromedial prefrontal lesions. Lesion studies such as these indicate the causal nature of the relationship between affect and moral judgment. Although much related work remains to be done, the existing body of evidence from multiple disciplines converges on the notion that moral judgments are not just correlated with emotion but result from emotion, as demanded by our account of the affect heuristic.

In addition, indirect evidence of attribute substitution comes from explanatory power. Attribute substitution can explain why common moral rules are defeasible, and also why it is so hard to specify when common moral rules are defeated. It is easy to say "don't break your promise unless there is an adequate

reason to break it,'' but it is hard to specify explicit rules that can be used to determine which promises may be broken. The affect heuristic view suggests that we consult our emotions or affect in order to determine when promises may be broken. That would explain why the rule against promise-breaking has exceptions and also why it is hard to specify what those exceptions are.

The heuristic hypothesis also explains why people seem able to answer complex moral questions surprisingly quickly, since attribute substitution would reduce reaction times—they are fast as well as frugal. It is hard to imagine any better explanation of such observations, given the inaccessibility of moral wrongness (discussed above, Section 2.1), so the best explanation seems to be that moral intuitions are attribute substitutions.

We still need evidence that this process is unconscious, but we already saw some. Both Hauser et al. (2008) and Mikhail (2002) found that subjects' moral judgments fit the doctrine of double effect, but very few subjects cited that doctrine when asked to justify their judgments. Additional evidence of unconsciousness comes from subjects' tendency to reject certain moral heuristics on reflection. When subjects become conscious that their moral judgments are affected by physical contact, for example, they often deny that contact really affects moral wrongness (Cushman et al., 2006). Since they reject the heuristic when asked to explain themselves, it seems unlikely that they were conscious of it at the earlier time when they formed their moral judgment. Subjects (and indeed many kinds of philosophers) also seem reluctant to endorse other heuristics, such as when they deny that their moral judgments are based on affect, so the same argument suggests that those other heuristics also operate unconsciously.

Finally, if moral heuristics are unconscious in the same way as non-moral heuristics, we would expect both to be partly but not wholly correctable by slower conscious reflection. In studies of the representativeness heuristic, subjects made fewer mistakes about Tom's field of study when they not only use the heuristic attribute of representativeness but also took time to consciously reflect on the additional information about the relative populations in various fields (Kahneman & Frederick, 2005). Analogously, when asked whether an act is morally wrong, people who initially use the affect heuristic can later correct their initial impressions if they take time to reflect on additional features, such as the consequences of actions. A neat example is Bentham's (1978) apparent disgust at gay sex while his utilitarian theory enables him to override that feeling and judge that gay sex is not morally wrong. See also Greene et al. (2004) on more recruitment of brain regions for abstract reasoning among those who choose to smother the crying baby. In such cases, moral heuristics seem

unconscious but consciously correctable in much the same way as non-moral heuristics.

Although moral intuitions resemble non-moral heuristics in all of these ways, there are differences as well. First, with one exception, none of the correlations between candidates for a heuristic attribute and moral wrongness is close to the 0.97 correlation that Kahneman et al. (1982) found with non-moral heuristics. However, the moral cases are more complex. Several different moral heuristics could operate on a single case. The availability of alternative heuristics, along with cultural overlays and other complications, should reduce correlations between moral wrongness and any one heuristic attribute in moral cases.

Moral heuristics are also more intractable than non-moral heuristics. When someone points out that every seven-letter word ending in "ing" has "n" in its sixth place, most people quickly admit that they made a mistake. People seem much more resistant to giving up their moral judgments even when they admit the analogies to heuristics. Why? One reason might be that they are not caught in an explicit inconsistency, as they are in the case of availability, so it is much harder to show them that their moral judgment is mistaken. People have a tendency to stick by their heuristics unless proven inconsistent or incorrect, which is harder to do in moral cases.[9] FN:9

Another reason for resistance is that giving up the heuristic has more costs in the moral case. People stick by their moral intuitions because they would feel uncomfortable without them, and because they know that others would be less comfortable with them if they did not share their moral views. None of that applies to cases like Tom the graduate student, the number of seven-letter words in a passage, or which German cities have soccer teams. In addition, some of our most valued social practices and institutions depend on using heuristics such as don't let your friends down. Without the tendency to stand by our friends even when doing so has bad consequences, our friendships would be very different and might not even count as friendships at all. That gives people extra incentive to resist giving up their heuristic-based moral judgments.[10] FN:10

Of course, numerous serious complications remain.[11] Many more kinds of moral intuitions need to be tested. Much more empirical work needs to be done. Nonetheless, this research project seems promising in psychology.[12] FN:11 FN:12

[9] In this way, moral target attributes are even less accessible than non-moral target attributes.

[10] This introduces a kind of partiality into moral intuitions insofar as those moral intuitions depend our personal interests and desires.

[11] Moral heuristics might contain several levels or be used in chains. We might usually follow conscious rules and turn to heuristics, such as the affect heuristic, only when rules run out. Some heuristics might be hardwired by biology, although culture still might affect which heuristics we use.

[12] One promise is to explain adolescents. Adolescents tend to use non-moral heuristics less than adults do, so the view of moral intuitions as heuristics suggests that adolescents would also not have

3. Some Philosophical Implications

We have not, of course, come close to showing that *all* moral intuitions fit the heuristic model. As we said at the start, we have been talking only about a subclass of moral intuitions—intuitions about specific cases. Only a small subclass of that subclass has been tested experimentally. Still, it is worthwhile to ask what would follow if the heuristic model turned out to hold for all moral intuitions. On that assumption, the heuristic model of moral intuitions might have several important philosophical implications.

3.1. *Direct Insight*

First, if moral intuitions result from heuristics, moral intuitionists (cf. Stratton-Lake, 2003) must stop claiming direct insight into moral properties. This claim would be as implausible as claiming direct insight into probability or numbers of seven-letter words, based on how we employ the representativeness and availability heuristics. Heuristics often *seem* like direct insight, but they never really *are* direct insight, because they substitute attributes. If moral judgments are reached through mediating emotions or affect, then they are not reached directly.

Some moral intuitionists might respond that they claim direct insight only *after* reflection (Audi, 2004). However, subjects also reflect when they use non-moral heuristics, like representativeness and availability, to estimate probabilities and numbers of words. Thus the presence of reflection does not show either that no heuristic is employed or that the insight is direct.

3.2. *Reliability*

Second, the view that moral intuitions result from heuristics raises doubts about whether and when we should trust moral intuitions. Just as non-moral heuristics lack reliability in unusual situations, so do moral intuitions, if they are based on moral heuristics. It would be interesting and important (though challenging) to do the empirical work needed to determine which moral intuitions are reliable in which circumstances, just as Gigerenzer and his colleagues are trying to do for non-moral heuristics.

Of course, just as disagreements abound over whether the use of heuristics is rational or not in the non-moral domain (see Gigerenzer & Kahneman),

fully-formed moral intuitions of some kinds. This would explain both crazy behavior by adolescents and also moral relativism among adolescents (cf. Nichols, 2004).

similar concerns will arise for moral intuitions. Which specific moral intuitions are products of heuristics? Are heuristic-driven intuitions especially unreliable? If so, in what circumstances? Should the mechanisms by which heuristics exert their effects inform our answers to the previous questions? For instance, if the affect heuristic turns out indeed to be the master heuristic (in place of or in addition to a set of heuristics concerning specific and apparently morally relevant acts, such as lying or killing), should we alter our attitude towards the resultant intuitions? All of these questions can stimulate new philosophical thinking about morality.

3.3. *Counter-Examples and Consequentialism*

Finally, this account of moral intuitions helps to defend consequentialism and other moral theories against simple counter-examples. Critics often argue that consequentialism can't be accurate, because it implies moral judgments that are counter-intuitive, such as that we are morally permitted to punish an innocent person in the well-known example where this is necessary to stop riots and prevent deaths. With the heuristic model in hand, consequentialists can respond that the target attribute is having the best consequences, and any intuitions to the contrary result from substituting a heuristic attribute.

Of course, consequentialists can't just assert these claims without support. They need to explain how these heuristic attributes are related to the target attribute in order to understand why we evolved with these heuristics rather than with others. But that's not too hard, because the usual heuristic attributes do seem to be good indicators of the best consequences in common circumstances (Baron, 1994). That is why rule-utilitarians can support common moral intuitions (Hooker, 2000).

If moral intuitions reflect normal consequences, and if that is why they evolved, then moral intuitions in unusual circumstances have no more force against consequentialism than folk intuitions about Tom have against probability theory. Consequentialists can go on to admit that moral intuitions are useful and even necessary as a practical guide or decision procedure, as long as the standard or criterion of right remains the target attribute of having the best consequences (Bales, 1971). This nice fit between consequentialism and the heuristic model of moral intuitions does not give any positive support for either, but it does help to defend consequentialism against common objections, and it shows that the heuristic model is relevant to normative issues.

Similar moves are also available for other moral theories, as long as they make moral wrongness inaccessible in practice and can explain why the heuristic attributes are good indicators of moral wrongness. Many moral theories can

do that, so the heuristic account of moral intuitions is neutral among those moral theories. The lesson is not about which moral theory is true but, instead, about which method to use in choosing among moral theories. The heuristic model suggests that moral philosophy cannot be done simply by means of counter-examples and accusations that opposing theories are counter-intuitive. Instead, we need to consider how those contrary moral intuitions arose, via which heuristic or set of heuristics. In this and other ways, the heuristic model of moral intuitions has important implications for methodology in moral philosophy.

References

Alicke, M. D. 1992. Culpable Causation. *Journal of Personality and Social Psychology, 63*: 368–78.

Audi, Robert. 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, NJ: Princeton University Press.

Bales, R. E. 1971. Act-utilitarianism: account of right-making characteristics or decision-making procedures? *American Philosophical Quarterly*, 8: 257–65.

Baron, J. 1994. Nonconsequentialist Decisions. *Behavioral and Brain Sciences, 17*: 1–10.

Bentham, J. 1978. Offences Against One's Self. *Journal of Homosexuality*, 3 (4): 389–405, and 4 (1): 91–107.

Brink, D. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge and New York: Cambridge University Press.

Carlsmith, K. M., Darley, J. M., and Robinson, P. H. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social* Psychology, *83:* 284–299.

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.

Ciaramelli, E., Muccioli, M., Làdavas, E., and di Pellegrino, G. 2007. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2 (2): 84–92.

Cushman, F., Young, L. and Hauser, M. 2006. The Role of Conscious Reasoning and Intuition in Moral Judgments: Testing Three Principles of Harm. *Psychological Science, 17*: 1082–1089.

Doris, J. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Gigerenzer, G. 2008. Moral Intuition = Fast and Frugal Heuristics? In *Moral Psychology, Volume 2: The Cognitive Science of Morality*, ed. W. Sinnott-Armstrong, pp. 1–26. Cambridge: MIT Press.

Gigerenzer, G., Todd, P., & the ABC Research Group. 1999. *Simple Heuristics that Make us Smart*. New York: Oxford University Press.

Greene, J. D., Lindsell, D., Clarke, A. C., Lowenberg, K., Nystrom, L. E., & Cohen, J.D. 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111 (3)*: 364–371.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107 (3)*: 1144–1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*: 389-400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science, 293*: 2105–2108.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*: 613–628.

Haidt, J., & Joseph, C. 2004. Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues. *Daedalus*: 55–66.

Harman, G. 1996. Moral Relativism. In *Moral Relativism and Moral Objectivity*, by G. Harman & J. J. Thomson, pp. 1–64. Oxford: Blackwell.

——1999. Moral philosophy meets social psychology. Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society 99*: 315–331.

Hauser, M., Young, L., & Cushman, F. 2008. Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Action. In *Moral Psychology, Volume 2: The Cognitive Science of Morality*, ed. W. Sinnott-Armstrong, pp. 107–143. Cambridge: MIT Press.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. New York: Oxford University Press.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.

Kahneman, D., & Frederick, S. 2005. A Model of Heuristic Judgment. In *The Cambridge Handbook of Thinking and Reasoning*, ed. K. J. Holyoak & R. G. Morrison, pp. 267–293. New York: Cambridge University Press.

Kahneman, D., Schkade, D., & Sunstein, C. R. 1998. Shared outrage and erratic rewards: The psychology of punitive damages. *Journal of Risk and Uncertainty, 16*: 49–86.

Kahneman, D., Slovic, P., & Tversky, A. 1982. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kass, Leon. 1997. The Wisdom of Repugnance. *The New Republic*, 2 June, 17–26.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. 2007. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, 446*: 908–911.

Laland, K. 2001. Imitation, Social Learning, and Preparedness as Mechanisms of Bounded Rationality. In *Bounded Rationality: The Adaptive Toolbox*, ed. G. Gigerenzer and R. Selten, pp. 233–248. Cambridge: MIT Press.

Mendez, M. F., Anderson, E., & Shapira, J. S. 2005. An Investigation of Moral Judgment in Frontotemporal Dementia. *Cognitive Behavioral Neurology, 18* (4): 193–197.

Merritt, M. 2000. Virtue ethics and situationist personality psychology. *Ethical Theory and Moral Practice, 3*: 365–383.

Mikhail, J. 2002. Aspects of a Theory of Moral Cognition: Investigating Intuitive Knowledge of the Prohibition of Intentional Battery and the Principle of Double Effect. Downloadable at http://ssrn.com/abstracts=762385

Nichols, Shaun. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.

Samuels, R., Stich, S., & Bishop, M. 2002. Ending the Rationality Wars: How to Make Disputes about Human Rationality Disappear. In *Common Sense, Reasoning and Rationality*, ed. R. Elio, pp. 236–268. New York: Oxford University Press.

Scanlon, T. M. 1999. *What We Owe to Each Other*. Cambridge, MA: Belknap Press.

Schaich Borg, J., Hynes, C., Grafton, S., & Sinnott-Armstrong, W. 2006. Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation. *Journal of Cognitive Neuroscience* 18 (5): 803–817.

Schnall, S., Haidt, J., Clore, G., & Jordan, A. 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34: 1096–1109.

Sinnott-Armstrong, W., Mallon, R., McCoy, T., & Hull, J. 2008. Intention, Temporal Order, and Moral Judgments. *Mind & Language* 23 (1): 90–106.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. 2002. The Affect Heuristic. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. T. Gilovich, D. Griffin, & D. Kahneman, pp. 397–420. New York: Cambridge University Press.

Stratton-Lake, P., ed. 2003. *Ethical Intuitionism: Re-evaluations*. New York: Oxford University Press.

Sunstein, Cass. 2005. Moral Heuristics. *Behavioral and Brain Sciences* 28: 531–73.

Valdesolo, P., & DeSteno, D. 2006. Manipulations of Emotional Context Shape Moral Judgments. *Psychological Science* 17 (6): 476–477.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16: 780–784.

Wilson, E. O. 1998. *Consilience: The Unity of Knowledge*. New York: Knopf.