Moral Heuristics

Cass R. Sunstein
CassR.Sunstein@chicagounbound.edu

Follow this and additional works at: https://chicagounbound.uchicago.edu/law_and_economics

Part of the Law Commons

Recommended Citation

This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.
Moral Heuristics

Cass R. Sunstein

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

March 2003

This paper can be downloaded without charge at:
and at the Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract_id=387941
Abstract

With respect to questions of fact, people use heuristics—mental short-cuts, or rules of thumb, that generally work well, but that also lead to systematic errors. People use moral heuristics too—moral short-cuts, or rules of thumb, that lead to mistaken and even absurd moral judgments. These judgments are highly relevant to law and politics. Examples are given from a number of domains, with an emphasis on appropriate punishment. Moral framing effects are discussed as well.

I. Introduction

Consider two positions voiced by many people in the American media in 2003:

1. When you have been a fan of a sports team, you have a moral obligation to continue to be a fan even if the team is now terrible. It is disloyal to cease being a fan merely because the team keeps losing. “Once you’re a fan, you’re a fan for life.”

2. In opposing military action to topple Saddam Hussein, France violated its moral obligations. The United States liberated France from Hitler’s Germany, and if the United States favored military action to topple Saddam Hussein, France was under a moral obligation to support the United States.

Both of these claims are absurd. A sports team is not a person or a friend, but a collection of strangers, whose personnel changes over time. There is nothing immoral about following and liking a team in one year, but ignoring or disliking that same team a few years later. You don’t have a moral obligation to continue to support a sports team. To be sure, France owes the United States a debt of gratitude and more. And if America’s position is right, France should agree with America. But France was not obliged to accept the American position on war with Iraq, certainly not if it believed that such a war would be unjust or against the interests of all those concerned. Of course there are vigorous arguments on behalf of collective responsibilities and group-level attachments. But even if we accept those arguments, and believe in collective responsibilities, we are most unlikely to be able to accept these claims about the moral obligations of sports fans and France.

* Karl N. Llewellyn Distinguished Service Professor of Jurisprudence, Law School and Department of Political Science, University of Chicago. I am grateful to Eric Posner and Adrian Vermeule for valuable comments on a previous draft.
What is interesting about these claims is not their absurdity, but the fact that both of them have a structure, one that makes them humanly recognizable rather than arbitrary or unintelligible. In both cases, people are overgeneralizing from a moral intuition that works well in daily life. In the case of the sports team, the moral intuition involves relationships between friends. It is morally wrong to be a “fair weather friend”; you shouldn’t abandon a friend who has fallen on hard times. If the relationship between fans and teams is analogous to the relationship between friends, then it is also morally wrong to be a “fair weather fan.” In the case of France, friendship is also the operative analogy. If people save your life, you should be loyal to them; you shouldn’t stab your benefactors in the back when they’re at risk. Both claims take sound moral intuitions, useful for most of life’s situations, and generalize them to superficially similar contexts in which those intuitions lose their foundations. In both cases, people are using moral heuristics—moral short-cuts, or rules of thumb, that work well most of the time, but that also systematically misfire.

My goal in this essay is to suggest that moral heuristics play a pervasive role in moral, political, and legal judgments, and that they produce serious mistakes. Often such heuristics represent generalizations from a range of problems for which they are indeed well-suited. Usually such heuristics work well. The problem comes when the generalizations are wrenches out of context and treated as freestanding or universal principles, applicable to situations in which their justifications no longer operate. Because the generalizations are treated as freestanding or universal, their application seems obvious, and those who reject them appear morally obtuse, possibly even monstrous. I want to urge that the appearance is misleading. And by drawing attention to moral heuristics, I hope to give a possible answer to the puzzling question about why people make persistent moral errors. Often, I suggest, the answer lies not in self-interest, stupidity, or venality, but in the use of heuristics that misfire.

In making these claims, I draw on existing psychological work on heuristics and biases, pioneered by Daniel Kahneman and Amos Tversky. That works deals not with moral questions, but with issues of fact. The basic claim here is that in answering hard factual questions, those who lack accurate information use simple rules of thumb. How many words, on a particular page, will have “ing” as the last three letters? How many words, on a particular page, will have the “n” as the second-to-last letter? Most people think that a significant number of words will end in “ing,” and that a smaller number of words will have “n” as the second-to-last letter—even though a moment’s reflection shows that this cannot possibly be true. People err because they use an identifiable

---

1 On overgeneralization, see Jonathan Baron, Nonconsequentialist Decisions, 17 Behavioral and Brain Sciences 1 (1994); I have learned a great deal from Baron’s analysis and from Baron’s work in general. See Jonathan Baron, Judgment Misguided: Intuition and Error in Public Decision Making (1998).

2 The key papers can be found in Judgment Under Uncertainty: Heuristics and Biases (Daniel Kahneman, Paul Slovic, and Amos Tversky eds., 1982); a more recent collection is Heuristics and Biases: The Psychology of Intuitive Judgment (Thomas Gilovich et al. eds. 2002). The heuristics-and-biases literature should be distinguished from the literature on prospect theory, which involves the nature of people’s utility functions under conditions of risk, not mental shortcuts under conditions of uncertainty. See Daniel Kahneman and Amos Tversky, Choices, Values, and Frames (2001).
heuristic—the availability heuristic—to answer difficult questions about probability. When people use this heuristic, they answer a question of probability by asking whether examples come readily to mind. How likely is a flood, an airplane crash, a traffic jam, a terrorist attack, or a disaster at a nuclear power plant? Lacking statistical knowledge, people try to think of illustrations. Thus, “a class whose instances are easily retrieved will appear more numerous than a class of equal frequency whose instances are less retrievable.” For people without statistical knowledge, it is far from irrational to use the availability heuristic; the problem is that this heuristic can lead to serious errors of fact, in the form of excessive fear of small risks and neglect of large ones.

Or consider the representativeness heuristic. The most famous example involves the likely career of a hypothetical woman named Linda, described as follows: “Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.” People were asked to rank, in order of probability, eight possible futures for Linda. Six of these were fillers (psychiatric social worker, elementary school teacher); the two crucial ones were “bank teller” and “bank teller and active in the feminist movement.” Most people said that Linda was less likely to be a bank teller than to be a bank teller and active in the feminist movement. This is an obvious logical mistake, a conjunction error, in which characteristics A and B are thought to be more likely than characteristic A alone. The error stems from the representativeness heuristic: Linda’s description seems to match “bank teller and active in the feminist movement” far better than “bank teller.” In an illuminating reflection on the example, Stephen Jay Gould observed that “I know [the right answer], yet a little homunculus in my head continues to jump up and down, shouting at me—‘but she can’t just be a bank teller; read the description.’”

A detailed literature has developed over the vices and virtues of the heuristics, some of them “fast and frugal,” that play a role in both life and law. But the relevant literature does not investigate the possibility that people might rely on simple rules of thumb, for purposes of assessing moral and political issues, that often work well but that often misfire. In fact the central point seems obvious. Much of everyday morality consists of simple rules that generally make sense but that fail in certain cases. It is wrong to lie or steal, but if a lie or a theft would save a human life, lying or stealing is probably obligatory. Not all promises should be kept. It is wrong to try to get out of a longstanding social commitment at the last minute, but if your child is in the hospital, you are morally required to do exactly that. And if good heuristics misfire in the factual domain, they will inevitably do so in the domains of morality and law as well.

---

3 See Tversky and Kahneman, supra note 2, at 3, 11-14.
4 Id. at 11.
7 See supra note; Gerd Gigerenzer et al., Simple Heuristics That Make Us Smart (1999).
I believe that an understanding of moral heuristics casts light on a number of widely held but ultimately implausible and sometimes even absurd intuitions—intuitions that belong in the same category as those involving baseball fans and France. These intuitions play a large role in both politics and law. An understanding of moral heuristics simultaneously raises doubts about certain methods of inquiry in moral and political philosophy, methods that depend on eliciting moral intuitions about exotic cases that people almost never face in daily life. Those intuitions, I suggest, are peculiarly unreliable when they are wrenching out the contexts in which they make sense. For example, people might be asked to consider whether they would kill an innocent person in order to save twenty people who would otherwise be killed. These questions, abstracted from real life, often produce firm moral judgments that are then treated, in the analysis, as important “data” for testing moral and political theories and for figuring out what really should be done.

I believe that there is a serious problem with this way of proceeding. Indeed, philosophers who proceed in this way seem to me inadvertently to be following the research agenda used by Kahneman and Tversky: Design a case, never actually encountered, in which intuitions, however firm, are likely to go wrong. Because Kahneman and Tversky were dealing with facts, they could demonstrate that the intuitions led to errors. Unfortunately, that cannot be demonstrated here. In the moral domain, it is hard to come up with unambiguous cases where the error is both highly intuitive and on reflection uncontroversial—where people can easily be embarrassed about their own intuitions (and despite the embarrassment, hear the continued squawking of their own version of Gould’s homunculus). But I think that the sports fan and France cases belong in just this domain and that other moral intuitions, of far greater practical importance, can be similarly understood. My minimal suggestions are that moral heuristics exist—and that it is odd to treat the resulting moral intuitions as fixed points for analysis, rather than as unreliable and at least potentially erroneous.

II. Moral Heuristics in Action

A. Excessively Ambitious Starts

If we take seriously the possibility that moral claims operate as heuristics, we might be able to imagine some very ambitious claims. Some of the largest moral theories might be characterized, and ultimately rejected, in this way. Consider the view that much of everyday morality, nominally concerned with fairness, should be seen as a set of heuristics for the real issue, which is how to promote utility. Armed with psychological findings about the use of heuristics, utilitarians might be tempted to claim that ordinary moral commitments are a set of mental shortcuts that generally work well, but that also

---

produce severe and systematic errors. Suppose people are committed to retributivism; this is their preferred theory of punishment. Might they be making a cognitive error? (Is Kantianism a series of cognitive errors?)

But easy victories are unlikely here. Indeed, those who reject utilitarianism might easily turn the tables. They might contend that the rules recommended by utilitarians are consistent, much of the time, with what morality requires -- but also that utilitarianism, taken seriously, produces bad mistakes in some cases. And indeed, many debates between utilitarians and their critics involve claims, by one or another side, that the opposing view leads to results, in particular cases, that are inconsistent with widespread intuitions and should be rejected for that reason. These large debates are unlikely to be tractable, simply because utilitarians and deontologists are most likely to be unconvinced by the suggestion that their defining commitments are mere heuristics. Here there is a large difference between moral heuristics and the heuristics uncovered in the relevant psychological work, where the facts, or simple logic, provide a good test whether people have erred. If people tend to think that more words, on a given page, end with the letters “ing” than have “n” in the next-to-last position, something has clearly gone wrong. If people think that some person Linda is more likely to be a “a bank teller who is active in the feminist movement” than a “feminist bankteller,” there is an evident problem. In the cases that concern me here, factual blunders and simple logic do not provide such a simple test.

But in some particular cases, we might be able to make some progress by entertaining the hypothesis that some widely accepted rules of morality are heuristics. Consider several possibilities.

B. Pointless Punishment

People’s intuitions about punishment seem quite disconnected with the consequences of punishment, in a way that suggests that a moral heuristic may well be at work. Suppose, for example, that a corporation has engaged in serious wrongdoing. People are likely to want to punish the corporation as if it were a person.11 They are unlikely to inquire into the possibility that the consequences of serious punishment (say, a stiff fine) will not be to “hurt” corporate wrongdoers, but instead to decrease wages, increase prices, or produce lost jobs. Punishment judgments are rooted in a simple heuristic, to the effect that penalties should be a proportional response to the outrageousness of the act. In thinking about punishment, people use an outrage heuristic.12 According to this heuristic, people’s punishment judgments are a product of their outrage. This heuristic produces sensible results in most circumstances, but in some cases, it seems to me to lead to systematic errors.

---

Consider, for example, an intriguing study of people’s judgments about penalties in cases involving harms from vaccines and birth control pills.\textsuperscript{13} In one case, subjects were told that the result of a higher penalty would be to make companies try harder to make safer products. In an adjacent case, subjects were told that the consequence of a higher penalty would be to make the company more likely to stop making the product, with the result that less safe products would be on the market. Most subjects, including a group of judges, gave the same penalties in both cases. Can this outcome be defended in principle? I doubt that it can. I think that it is far more sensible to think that people are operating under a heuristic, requiring punishment that is proportional to outrageousness, and should not be based on consequential considerations. As a general rule, of course, it is plausible to think that penalties should be proportional to the outrageousness of the act. But it is fanatical to insist on this principle whether or not the consequence would be to make human beings safer and healthier. Those who insist on proportional punishments might disagree, but it might be worthwhile for them to consider the possibility that they have been tricked by a heuristic.

If this claim seems too adventurous, consider a similar test of punishment judgments, which asked subjects, including judges and legislators, to choose penalties for dumping hazardous waste.\textsuperscript{14} In one case, the penalty would make companies try harder to avoid waste. In another, the penalty would lead companies to cease making a beneficial product. Most people did not penalize companies differently in the two cases. Most strikingly, people preferred to require companies to clean up their own waste, even if the waste did not threaten anyone, instead of spending the same amount to clean up far more dangerous waste produced by another, now-defunct company. How could this preference make sense? Why should a company be asked to engage in a course of action that costs the same but that does much less good? I believe that people are using a heuristic, essentially requiring people to correct their own wrongs, even in a case in which that heuristic leads to palpably inferior results.

C. Betrayals

To say the least, people do not like to be betrayed. A betrayal of trust is likely to produce a great deal of outrage. If a babysitter neglects a child, or if a security guard steals from his employer, people will be angrier than if the identical acts are performed by someone in whom trust has not been reposed. So far, perhaps, so good. And it should not be surprising that people will favor greater punishment for betrayals than for otherwise identical crimes.\textsuperscript{15} Perhaps the disparity can be justified on the ground that the

\textsuperscript{13} Jonathan Baron and Ilana Ritov, Intuitions About Penalties and Compensation in the Context of Tort Law, 7 J Risk and Uncertainty 17 (1993)


\textsuperscript{15} See Jonathan J. Koehler and Andrew D. Gershoff, Betrayal Aversion: When Agents of Protection Become Agents of Harm 89 Organizational Beh and Human Dec Processes (2002) (forthcoming) (relying on five empirical studies to find that acts of betrayal elicited stronger desired punishments than other bad acts).

\textsuperscript{15} Id at 40 (finding that “when faced with a choice among pairs of safety devices… most people preferred inferior options (in terms of risk exposure) to options that included a slim (0.01%) risk of betrayal”).

6
betrayal of trust is an independent harm, one that warrants greater deterrence and retribution—a point that draws strength from the fact that trust, once lost, is not easily regained. A family robbed by its babysitter is more seriously injured than a family who has been robbed by a thief. The loss of money is compounded and possibly dwarfed by the violation of a trusting relationship. The consequence of the violation might also be more serious. Will the family ever feel entirely comfortable with babysitters? It is bad to have an unfaithful spouse, but it is even worse if the infidelity occurred with your best friend.

In this light it is possible to understand why betrayals produce special moral opprobrium and (where the law has been violated) increased punishment. But consider a finding that is much harder to explain: People are especially averse to risks of death that come from products (like airbags) designed to promote safety. The aversion is so great that people have been found to prefer a chance of dying, as a result of accidents from a crash, to a significantly lower chance of dying in a crash as a result of a malfunctioning airbag. In other words, people’s aversion to betrayals is so great that they will increase their own risks rather than subject themselves to a (small) hazard that comes from a device that is supposed to increase safety. Indeed, “most people are willing to double their chance of dying to avoid incurring a very small chance of dying via betrayal.”

What explains this seemingly bizarre and self-destructive preference? I suggest that a heuristic is at work: Punish betrayals of trust. The heuristic generally works well. But it misfires crazily in some cases, as when those who deploy it end up increasing the risks they themselves face. An airbag is not a security guard or a babysitter, endangering those whom they have been hired to protect. It is a product, to be chosen if and only if it decreases aggregate risks. If an airbag makes people safer on balance, it should be used, even if in a tiny percentage of cases it will create a risk that would not otherwise exist. To reject airbags on grounds of betrayal is not entirely rational—not entirely rational but understandable, the sort of mistake to which heuristics often lead human beings. People’s unwillingness to subject themselves to betrayal risks, in circumstances in which products are involved and they are increasing their likelihood of death, is the moral cousin to the use of the representativeness heuristic in the Linda case.

D. Cost-Benefit Analysis

An automobile company is deciding whether to take certain safety precautions for its cars. In deciding whether to do so, it conducts a cost-benefit analysis, in which it concludes that certain precautions are not justified—because, say, they would cost $100 million and save only four lives, and because the company has a “ceiling” of $10 million per lives saved (a ceiling that is, by the way, significantly higher than the amount the Environmental Protection Agency uses for a statistical life). How will ordinary people react to this decision? The answer is that they will not react favorably. In fact they tend to punish companies that base their decisions on cost-benefit analysis, even if a high

16 Id.
17 Id.
valuation is placed on human life. By contrast, they do not much punish companies that are willing to impose a “risk” on people.19 What underlies these moral judgments?

A careful look raises the possibility that when people disapprove of trading money for risks, they are generalizing from a set of moral principles that are generally sound, and even useful, but that work poorly in some cases. Consider the following moral principle: Do not knowingly cause a human death. People disapprove of companies that fail to improve safety when they are fully aware that deaths will result—whereas people do not disapprove of those who fail to improve safety while appearing not to know, for certain, that deaths will ensue. When people object to risky action taken after cost-benefit analysis, it seems to be partly because that very analysis puts the number of expected deaths squarely “on screen.”20 Companies that fail to do such analysis, but that are aware that a “risk” exists, do not make clear, to themselves or to anyone else, that they caused deaths with full knowledge that this was what they were going to do. People disapprove, above all, of companies that cause death knowingly.

I suggest, then, that a moral heuristic is at work, one that imposes moral condemnation on those who knowingly engage in acts that will result in human deaths. And of course this heuristic does a great deal of good.21 The problem is that it is not always unacceptable to cause death knowingly, at least if the deaths are relatively few and an unintended byproduct of generally desirable activity. When government allows new highways to be built, it knows that people will die on those highways; when government allows new power plants to be built, it knows that some people will die from the resulting pollution; when companies produce tobacco products, and when government does not ban those products, hundreds of thousands of people will die; the same is true for alcohol. Much of what is done, by both industry and government, is likely to result in one or more deaths. Of course it would make sense, in most or all of these domains, to take extra steps to reduce risks. But that proposition does not support the implausible claim that we should disapprove, from the moral point of view, of any action taken when deaths are foreseeable.

I do believe that it is impossible to vindicate, in principle, the widespread social antipathy to cost-benefit balancing. But to adapt Stephen Jay Gould’s claim about the representativeness heuristic, “a little homunculus in my head continues to jump up and down, shouting at me” that corporate cost-benefit analysis, trading dollars for a known number of deaths, is morally unacceptable. The voice of the homunculus, I am suggesting, is not reflective, but instead a product of a crude but quite tenacious moral heuristic.22

19 See id. See also See Philip Tetlock, Coping With Tradeoffs, in Elements of Reason: Cognition, Choice, and the Bounds of Rationality 239, Arthur Lupia et al. eds. (Cambridge: Cambridge University Press, 2000)
20 It is also the case that explicit trading of money for lives is strongly disfavored, see Tetlock, supra note. I am hypothesizing that some of this effect, and possibly a great deal of it, comes from the fact that someone has knowingly engaged in action that will result in deaths.
21 It is even possible that the heuristic leads to better results, on balance, than a more refined approach that attempts to distinguish among different situations. See the discussion below.
E. Acts and Omissions

To say the least, there has been much discussion of whether and why the distinction between acts and omissions might matter for morality, law, and policy. In one case, for example, a patient might ask a doctor not to provide life-sustaining equipment, thus ensuring the patient’s death. In another case, a patient might ask a doctor to inject a substance that will immediately end the patient’s life. Many people seem to have a strong moral intuition that the failure to provide life-sustaining equipment, and even the withdrawal of such equipment, is acceptable and legitimate -- but that the injection is morally abhorrent. And indeed American constitutional law reflects judgments to this effect.22 But what is the morally relevant difference?

It is worth considering the possibility that the action-omission distinction operates as a heuristic for the more complex and difficult assessment of the moral issues at stake. From the moral point of view, harmful acts are generally worse than harmful omissions, in terms of both the state of mind of the wrongdoer and the likely consequences of the wrong. A murderer is typically more malicious than a bystander who refuses to come to the aid of someone who is drowning; the murderer wants his victim to die, whereas the bystander need have no such desire. In addition, a murderer typically guarantees death, whereas many bystanders do no such thing. But in terms of either the wrongdoer’s state of mind or the consequences, harmful acts are not always worse than harmful omissions. The moral puzzles arise when life, or a clever interlocutor, comes up with a case in which there is no morally relevant distinction between acts and omissions, but when moral intuitions (and the homunculus) strongly suggest that there must be such a difference. In such cases, we might hypothesize that moral intuitions reflect an overgeneralization of principles that usually make sense—but that fail to make sense in the particular case.23 Those principles condemn actions but permit omissions, a difference that is often plausible in light of relevant factors but that, in hard cases, cannot be defended. I believe that the persistent acceptance of withdrawal of life-saving equipment, alongside persistent doubts about euthanasia, is a demonstration of the point.

Consider in this regard the dispute over two well-known problems in moral philosophy.24 The first, called the trolley problem, asks people to suppose that a runaway trolley is headed for five people, who will be killed if the trolley continues on its current course. The question is whether you would throw a switch that would move the trolley onto another set of tracks, killing one person rather than five. Most people would throw the switch. The second, called the footbridge problem, is the same as that just given, but with one difference: the only way to save the five is to throw a stranger, now on a footbridge that spans the tracks, into the path of the trolley, killing that stranger but preventing the trolley from reaching the others. Most people will not kill the stranger. But what is the difference between the two cases, if any? A great deal of philosophical work has been done on this question, much of it trying to suggest that our firm intuitions can

indeed be defended in principle.\(^{25}\) Let me suggest a simpler answer. As a matter of principle, there is no difference between the two cases. People’s different reactions are based on moral heuristics that condemn the throwing of the stranger but support the throwing of the switch. These heuristics generally point in the right direction. But they misfire in drawing a distinction between the two cleverly devised cases. What makes the cases difficult is that they are not distinguishable in principle, but moral heuristics, rooted in quite different situations, categorize them as very different. Hence people struggle heroically to rescue our intuitions and to establish that the two cases are genuinely different in principle. But they aren’t. In this sense, the action-omission distinction leads to errors.

Is there anything to be said to those who believe that their moral judgments, distinguishing the trolley and footbridge problems, are entirely reflective, and reflect no heuristic at all? Consider a suggestive experiment, designed to see how the human brain responds to the two problems.\(^{26}\) The authors do not attempt to answer the moral questions in principle, but they find “that there are systematic variations in the engagement of emotions in moral judgment,\(^{27}\) and that brain areas associated with emotion are far more active in contemplating the footbridge problem than in contemplating the trolley problem.\(^{28}\) Of course this experiment is not decisive; there may be good moral reasons why certain brain areas are activated by one problem and not by the other. Perhaps the brain is closely attuned to morally irrelevant difference. But as in the case of fear, where an identifiable region of the brain makes helpfully immediate but not entirely reliable judgments,\(^{29}\) and where other, also identifiable regions can supply correctives, so too, I believe, in the context of morality, politics, and law.

F. Probability of Detection

Now turn to a final example from the domain of punishment. On the economic account, the state’s goal, when imposing penalties for misconduct, is to ensure optimal deterrence.\(^{30}\) To increase deterrence, the law might increase the severity of punishment, or instead increase the likelihood of punishment. A government that lacks substantial enforcement resources might impose high penalties, thinking that it will produce the right deterrent “signal” in light of the fact that many people will escape punishment altogether. A government that has sufficient resources might impose a lower penalty, but enforce the law against all or almost all violators. These ideas lead to a simple theory in the context of punitive damages for wrongdoing: The purpose of such damages is to make up for the shortfall in enforcement. If injured people are 100\% likely to receive compensation, there is no need for punitive damages. If injured people are 50\% likely to receive compensation, those who bring suit should receive a punitive award that is twice the

\(^{25}\) See Thomson, supra note.
\(^{27}\) Id. at 2106.
\(^{28}\) Id.
amount of the compensatory award. The simple exercise in multiplication will ensure optimal deterrence.

But there is a large question whether people accept this account, and if not, why not. (For the moment, let us point to one side the question whether they should accept it in principle.) Experiments suggest that people reject optimal deterrence and that they do not believe that the probability of detection is relevant to punishment. The reason is that they use the outrage heuristic. I participated in two experiments designed to cast light on this question. In the first, we gave people cases of wrongdoing, arguably calling for punitive damages, and also provided people with explicit information about the probability of detection. Different people saw the same case, with only one difference: varying probability of detection. People were asked about the amount of punitive damages that they would choose to award. Our goal was to see if people would impose higher punishments when the probability of detection was low. In the second experiment, we asked people to evaluate judicial and executive decisions to reduce penalties when the probability of detection was high, and to increase penalties when the probability of detection was low. We wanted people to say whether they approved or disapproved of varying the penalty with the probability of detection.

Our findings were simple and straightforward. The first experiment found that varying the probability of detection had no effect on punitive awards. Even when people’s attention was explicitly directed to the probability of detection, people were indifferent to it. People’s decisions about appropriate punishment were unaffected by seeing a high or low probability of detection. The second experiment found that strong majorities of respondents rejected judicial decisions to reduce penalties because of high probability of detection -- and also rejected executive decisions to increase penalties because of low probability of detection. In other words, people did not approve of an approach to punishment that would make the level of punishment vary with the probability of detection. What apparently concerned them was the extent of the wrongdoing and the right degree of moral outrage—not optimal deterrence.

Of the various problems I have discussed, the rejection of optimal deterrence is the most difficult for my claim that people rely on moral heuristics that generally work well but that systematically misfire. The rejection is difficult for my claim because many people have principled reasons for rejecting that account of punishment (and I personally do not believe in it, at least not as a complete theory of punishment). Most people are intuitive retributivists of one or another kind, thinking that companies should be punished in proportion to their wrongdoing. In their view, it is absurd to improve severe punishment on a company merely because the probability of detection was low—unless it could be shown that the company was especially sneaky or stealthy (and hence that its conduct was especially outrageous). In the same vein, it might seem odd to be unusually lenient to a company whose wrongdoing would inevitably be detected and punished.

31 See Kahneman and Frederick, supra note.
To say the least, I do not mean to offer a final judgment on retributivism here. But it seems implausible to suggest that the aggregate level of misconduct is entirely irrelevant to punishment, or to ignore the fact that a system that refuses to impose enhanced punishment on hard-to-detect wrongdoing will end up with a great deal of wrongdoing. Even if retribution is an important or dominant part of a good system of punishment, consequences matter as well. Surely steps should be taken, other things being equal, to deter acknowledged wrongdoing, and to impose extra deterrence in cases where it is needed. People’s unwillingness to take any account of the probability of detection suggests that a moral heuristic is at work, one that leads to real errors.

III. Beyond Heuristics

A. Exotic Cases and Moral Judgments

Some of these examples will seem more controversial than others. But taken as a while, they seem to me to raise serious doubts about the wide range of work that approaches moral and political dilemmas by attempting to uncover moral intuitions about exotic cases of the kind never or rarely encountered in ordinary life. Should you shoot an innocent person if that is the only way to save twenty innocent people? What is the appropriate moral evaluation of a case in which a woman accidentally puts cleaning fluid in her coffee, and her husband, wanting her dead, does not provide the antidote, which he happens to have handy?

I believe that the underlying moral intuitions ordinarily work well, but that when they are wrenched out of familiar contexts, their reliability, for purposes of moral and legal analysis, is unclear. Consider the following intuition: Do not kill an innocent person, even if this is necessary to save others. (I put to one side the context of war.) In all likelihood, a society does much better if most people have this intuition, if only because judgments about necessity are likely to be unreliable and self-serving. But in a hypothetical case, in which it really is necessary to kill an innocent person to save twenty others, our intuitions might well turn to be unclear and contested -- and if our intuitions about the hypothetical case turn out to be very firm (do not kill innocent people, ever!), they might not deserve to be so firm, simply because they have been wrenched out of the real world context, which is where they need to be to make sense.

In short, I believe that some legal and philosophical analysis, based on exotic moral dilemmas, is replicating the early work of Kahneman and Tversky: uncovering situations in which intuitions, normally quite sensible, turn out to misfire. The irony is that while Kahneman and Tversky meant to devise cases that would demonstrate the misfiring, some philosophers devise cases with the thought that the intuitions are reliable and should form the building blocks for sound moral judgments. An understanding of how heuristics works suggests reason to doubt the reliability of those intuitions, even when they are very firm.

33 Cf. Williams, supra note.
34 See Thomson, supra note, at 31.
Now it is possible that the firmness of the underlying intuitions is actually desirable. Perhaps social life is better, not worse, because of the large number of people who treat heuristics as moral rules and who believe (for example) that innocent people should never be killed. If the heuristic is treated as a freestanding principle, perhaps some mistakes will be made, but only in exotic and rare cases, and perhaps people who accept the principle will avoid the temptation to depart from it when the justification appears sufficient but really isn’t. In other words, a firm rule might misfire in some cases, but it might be better than a more fine-grained approach, which, in practice, would misfire even more. (Those who believe that you should always tell the truth may do and be much better, all things considered, than those who believe that truth should be told only on the basis of case-specific, all-things-considered judgments in its favor. Indeed, those who stick with their sports teams, out of a misplaced sense of moral obligation, may well be more loyal friends than those who feel no such sense.) My suggestion is not that the moral heuristics, in their most rigid forms, are socially worse than the reasonable alternatives. It is hard to resolve that question in the abstract. I am claiming only that such heuristics lead to significant errors and a great deal of confusion.

B. Moral Framing

In cognitive psychology and behavioral economics, the study of heuristics has been part and parcel of a research program that has involved framing effects as well. Are there moral framing effects? If the answer is affirmative, we would have further reason to believe in moral heuristics, because the two are a product of related cognitive processes.

For a simple example of framing, consider the question whether to undergo a risky medical procedure. When people are told, “Of those who have this procedure, 90 percent are alive after five years,” they are far more likely to agree to the procedure than when they are told, “Of those who have this procedure, 10 percent are dead after five years.” Experience might be expected to solve this problem, but doctors too are vulnerable to this framing effect. Here the question does not involve moral and political issues. But a similar effect has been demonstrated in the important context of obligations to future generations, a much-disputed question of morality, politics, and law. A regulatory system that attempts to track people’s preferences would try to measure intergenerational time preferences, that is, to elicit people’s judgments about how to trade off the protection of current lives and future lives. In any case, an important question, asked in many debates about the issue, involves the nature of people’s moral judgments on that issue. And indeed, an influential set of studies finds that people believe that it is morally appropriate to prefer few lives in the current generation to many lives of those in

---

36 Id.
39 See Revesz, supra note.
future generations.40 From a series of surveys, Maureen Cropper and her coauthors suggest that people are indifferent between savings one life today and saving 45 lives in 100 years.41 They make this suggestion on the basis of questions asking people whether they would choose a program that saves “100 lives now” or a program that saves a substantially larger number “100 years from now.”42

But it turns out that other ways of framing the same problem yield radically different results.43 For example, most people consider “equally bad” a single death from pollution next year and a single death from pollution in 100 years44 -- implying no preference for members of the current generation. In a similar finding of no strong preference for the current generation, people are equally divided between two programs: one that will save 55 lives now and 105 more lives in twenty years; and one that will save 100 lives now and 50 lives 25 years from now.45 It is even possible to frame the question in such a way as to find that future lives are valued more, not less, highly than current lives.46 In short, people’s moral judgments about obligations to future generations are very much a product of framing effects. Undoubtedly the same effects can be found in many other domains.

These are speculative remarks on some complex subjects. If the use of heuristics is harder to demonstrate in the domain of morality than in the domain of facts, it is largely because we are able to agree, in the relevant cases, about what constitutes factual error, and often less able to agree about what constitutes moral error. But it is overwhelmingly likely that rules of thumb, generally sensible but also likely to go wrong, play a role not merely in factual judgments, but in the domains of morality, politics, and law as well.

Readers with comments should address them to:

Cass R. Sunstein
University of Chicago Law School
1111 East 60th Street
Chicago, IL  60637
csunstei@midway.uchicago.edu

41 Id.
42 Id.
43 Frederick, supra note.
44 Id. at 43.
45 Id. at 44.
46 Id. at 45. Frederick asked subjects to choose between two programs. The first would become more effective over time, saving 100 lives this decade, 200 lives in the following decade, and 300 lives in the decade after that. The second would become less effective over time, saving 300 lives this decade, 200 lives in the following decade, and 100 lives in the decade after that. Most people preferred the first program, apparently suggesting that future lives are valued more highly. Id.
Chicago Working Papers in Law and Economics
(Second Series)

13. J. Mark Ramseyer, Credibly Committing to Efficiency Wages: Cotton Spinning Cartels in Imperial Japan (March 1993)
16. Lucian Arye Bebchuk and Randal C. Picker, Bankruptcy Rules, Managerial Entrenchment, and Firm-Specific Human Capital (August 1993)
17. J. Mark Ramseyer, Explicit Reasons for Implicit Contracts: The Legal Logic to the Japanese Main Bank System (August 1993)
20. Alan O. Sykes, An Introduction to Regression Analysis (October 1993)
22. Randal C. Picker, An Introduction to Game Theory and the Law (June 1994)
29. Daniel Shaviro, Budget Deficits and the Intergenerational Distribution of Lifetime Consumption (January 1995)
34. J. Mark Ramseyer, Public Choice (November 1995)
41. John R. Lott, Jr. and David B. Mustard, Crime, Deterrence, and Right-to-Carry Concealed Handguns (August 1996)
42. Cass R. Sunstein, Health-Health Tradeoffs (September 1996)
47. John R. Lott, Jr. and Kermit Daniel, Term Limits and Electoral Competitiveness: Evidence from California's State Legislative Races (May 1997)
48. Randal C. Picker, Simple Games in a Complex World: A Generative Approach to the Adoption of Norms (June 1997)
50. Cass R. Sunstein, Daniel Kahneman, and David Schkade, Assessing Punitive Damages (with Notes on Cognition and Valuation in Law) (December 1997)
52. John R. Lott, Jr., A Simple Explanation for Why Campaign Expenditures are Increasing: The Government is Getting Bigger (February 1998)
| 60. | John R. Lott, Jr., How Dramatically Did Women's Suffrage Change the Size and Scope of Government? (September 1998) |
| 64. | John R. Lott, Jr., Public Schooling, Indoctrination, and Totalitarianism (December 1998) |
| 67. | Yannis Bakos, Erik Brynjolfsson, Douglas Lichtman, Shared Information Goods (February 1999) |
| 68. | Kenneth W. Dam, Intellectual Property and the Academic Enterprise (February 1999) |
| 70. | Cass R. Sunstein, Must Formalism Be Defended Empirically? (March 1999) |
| 71. | Jonathan M. Karpoff, John R. Lott, Jr., and Graeme Rankine, Environmental Violations, Legal Penalties, and Reputation Costs (March 1999) |
75. Richard A. Epstein, Deconstructing Privacy: and Putting It Back Together Again (May 1999)
76. William M. Landes, Winning the Art Lottery: The Economic Returns to the Ganz Collection (May 1999)
77. Cass R. Sunstein, David Schkade, and Daniel Kahneman, Do People Want Optimal Deterrence? (June 1999)
78. Tomas J. Philipson and Richard A. Posner, The Long-Run Growth in Obesity as a Function of Technological Change (June 1999)
79. David A. Weisbach, Ironing Out the Flat Tax (August 1999)
81. David Schkade, Cass R. Sunstein, and Daniel Kahneman, Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages (September 1999)
82. Cass R. Sunstein, Nondelegation Canons (September 1999)
83. Richard A. Posner, The Theory and Practice of Citations Analysis, with Special Reference to Law and Economics (September 1999)
84. Randal C. Picker, Regulating Network Industries: A Look at Intel (October 1999)
90. David A. Weisbach, Should the Tax Law Require Current Accrual of Interest on Derivative Financial Instruments? (December 1999)
95. David Schkade, Cass R. Sunstein, Daniel Kahneman, Deliberating about Dollars: The Severity Shift (February 2000)
105. Jack Goldsmith and Alan Sykes, The Dormant Commerce Clause and the Internet (November 2000)
110. Saul Levmore, Conjunction and Aggregation (December 2000)
111. Saul Levmore, Puzzling Stock Options and Compensation Norms (December 2000)
112. Richard A. Epstein and Alan O. Sykes, The Assault on Managed Care: Vicarious Liability, Class Actions and the Patient's Bill of Rights (December 2000)
114. Cass R. Sunstein, Switching the Default Rule (January 2001)
116. Jack Goldsmith, Statutory Foreign Affairs Preemption (February 2001)
118. Cass R. Sunstein, Academic Fads and Fashions (with Special Reference to Law) (March 2001)
122. David A. Weisbach, Ten Truths about Tax Shelters (May 2001)
126. Douglas G. Baird and Edward R. Morrison, Bankruptcy Decision Making (June 2001)
127. Cass R. Sunstein, Regulating Risks after ATA (June 2001)
129. Richard A. Epstein, In and Out of Public Solution: The Hidden Perils of Property Transfer (July 2001)
130. Randal C. Picker, Pursuing a Remedy in Microsoft: The Declining Need for Centralized Coordination in a Networked World (July 2001)
131. Cass R. Sunstein, Daniel Kahneman, David Schkade, and Ilana Ritov, Predictably Incoherent Judgments (July 2001)
133. Lisa Bernstein, Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms, and Institutions (August 2001)
137. Eric A. Posner and George G. Triantis, Covenants Not to Compete from an Incomplete Contracts Perspective (September 2001)
139. Randall S. Kroszner and Philip E. Strahan, Throwing Good Money after Bad? Board Connections and Conflicts in Bank Lending (December 2001)
140. Alan O. Sykes, TRIPs, Pharmaceuticals, Developing Countries, and the Doha “Solution” (February 2002)
141. Edna Ullmann-Margalit and Cass R. Sunstein, Inequality and Indignation (February 2002)
145. David A. Weisbach, Thinking Ouside the Little Boxes (March 2002, Texas Law Review)
149. Cass R. Sunstein, Beyond the Precautionary Principle (April 2002)
152. Richard A. Epstein, Steady the Course: Property Rights in Genetic Material (May 2002)
156. Cass R. Sunstein and Adrian Vermeule, Interpretation and Institutions (July 2002)
159. Randal C. Picker, From Edison to the Broadcast Flag: Mechanisms of Consent and Refusal and the Propertization of Copyright (September 2002)
162. Lior Jacob Strahilevitz, Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks (September 2002)
163. David A. Weisbach, Does the X-Tax Mark the Spot? (September 2002)
164. Cass R. Sunstein, Conformity and Dissent (September 2002)
166. Douglas Lichtman, Uncertainty and the Standard for Preliminary Relief (October 2002)
171. Richard A. Epstein, Animals as Objects, or Subjects, of Rights (December 2002)
172. David A. Weisbach, Taxation and Risk-Taking with Multiple Tax Rates (December 2002)