

2003

Moral Heuristics and Moral Framing Lecture

Cass R. Sunstein

Follow this and additional works at: http://chicagounbound.uchicago.edu/journal_articles



Part of the [Law Commons](#)

Recommended Citation

Cass R. Sunstein, "Moral Heuristics and Moral Framing Lecture," 88 Minnesota Law Review 1556 (2003).

This Article is brought to you for free and open access by the Faculty Scholarship at Chicago Unbound. It has been accepted for inclusion in Journal Articles by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

Lecture

Moral Heuristics and Moral Framing^{*}

Cass R. Sunstein[†]

INTRODUCTION

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

- A. 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."
- B. 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

^{*} This essay is based on the 2003–04 John Dewey Lecture in the Philosophy of Law, delivered by Professor Sunstein at the University of Minnesota Law School on September 24, 2003. The lectureship is named in honor of John Dewey, American philosopher, educator, and scholar. The John Dewey Lectureship is funded by a grant from the John Dewey Foundation and is sponsored by the University of Minnesota Law School to provide a forum for significant scholarly contributions to the development of jurisprudence.

[†] Karl N. Llewellyn Distinguished Service Professor of Jurisprudence, Law School and Department of Political Science, University of Chicago. I am grateful to Jonathan Baron, Emily Buss, Mary Anne Case, Elizabeth Emens, Carolyn Frantz, Charles Larmore, Eric Posner, Adrian Vermeule, and John Yoo for valuable comments on a previous draft; I also thank participants in an extremely helpful work-in-progress lunch at the University of Chicago Law School. This essay is a revised and expanded version of the John Dewey Lecture, delivered at the University of Minnesota Law School in September 2003. I am grateful to the faculty and students at the law school for their kindness and superb questions and comments at the time; those questions and comments resulted in many revisions and improvements.

person or a friend, but a collection of strangers, whose personnel changes over time. There is nothing immoral about following and supporting a team in one year, but ignoring or disliking that same team a few years later. You do not 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 in return for America's acts during World War II—certainly not if France 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 endorse these claims about the moral obligations of sports fans and France.

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.¹ 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 should not 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 should not stab your benefactors in the back when they are 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

1. On overgeneralization, see Jonathan Baron, *Nonconsequentialist Decisions*, 17 BEHAV. & BRAIN SCI. 1, 4–5 (1994) [hereinafter Baron, *Nonconsequentialist Decisions*]. I have learned a great deal from Baron's analysis and from his work in general. Overgeneralization might be said to be distinguishable from misapplication of a principle, as when a principle, sensibly designed for one setting, is senselessly applied to another. See JONATHAN BARON, JUDGMENT MISGUIDED: INTUITION AND ERROR IN PUBLIC DECISION MAKING 8–20 (1998) [hereinafter BARON, JUDGMENT MISGUIDED]. By overgeneralization, I mean simply to refer to use of a principle that ordinarily works well in a situation in which it does not; I do not intend a distinction from the overlapping idea of misapplication.

foundations. In both cases, people are using *moral heuristics*—moral shortcuts, or rules of thumb, that work well most of the time, but that also systematically misfire.

My major 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.² Often such heuristics represent generalizations from a range of problems for which they are indeed well suited. Usually such heuristics work well. The problem arises when the generalizations are wrenched 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. By drawing attention to moral heuristics, I hope to give a possible answer to the puzzling question about why people persist in making moral errors. Often, I suggest, the answer lies not in self-interest, stupidity, or venality, but in the use of heuristics that misfire.

Pioneering the modern work on the use of heuristics in cognition, Daniel Kahneman and Amos Tversky contended that in assessing probabilities, “people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations.”³ This contention has produced an extensive literature,⁴ and much controversy 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.⁵

The relevant literature, however, rarely investigates the possibility that in the moral and political domain, people also rely on simple rules of thumb that often work well but that

2. I build throughout on a brief discussion in Cass R. Sunstein, *Hazardous Heuristics*, 70 U. CHI. L. REV. 751, 751–52 (2003).

3. Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 3 (Daniel Kahneman et al. eds., 1982) [hereinafter JUDGMENT UNDER UNCERTAINTY].

4. For a recent collection, see HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT (Thomas Gilovich et al. eds., 2002) [hereinafter HEURISTICS AND BIASES].

5. See *id.*; GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART 27–28 (1999).

sometimes misfire.⁶ 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 long-standing social commitment at the last minute, but if your child is in the hospital, you are morally required to do exactly that. In politics, party affiliation, political identification, and leaders operate as heuristics. A self-identified Democrat might favor an increase in the minimum wage simply because the party favors such increases; a self-identified Republican might have supported the war with Iraq simply because President Bush did so. In the same vein, many people used party affiliation as a heuristic for answering the question of whether President Clinton should have been impeached.

Similar processes are at work in law. Some people believe that if Justice Antonin Scalia supports a certain position, it is likely to be right; Justice Scalia's view can do the work of a more fine-grained inquiry into the issues at stake. For many Justices, there is a corresponding heuristic. The "Scalia heuristic" is supplemented by the "Breyer heuristic" and the "O'Connor heuristic"; there are "Warren Court" and "Lochner Court" heuristics too. Those who follow these heuristics much simplify their own processes of evaluation, but they might well end up adopting some conclusions that they would repudiate after more sustained reflection. Every law professor knows that a class discussion can be reframed, and students polarized, by specifically attributing a view to the Rehnquist Court or to Justice Brennan. For questions of policy, heuristics play a large role. Should human cloning be banned? Should pesticides be regulated more severely than organic foods? Answers to these questions inevitably reflect the influence of heuristics. And if good heuristics misfire in the factual domain, they will inevitably do so in the domains of morality, politics, and law as well.

I believe that an understanding of moral, political, and legal 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

6. Exceptions include BARON, JUDGMENT MISGUIDED, *supra* note 1, and David Messick, *Equality as a Decision Heuristic*, in PSYCHOLOGICAL PERSPECTIVES ON JUSTICE (Barbara Mellers & Jonathan Baron eds., 1993).

baseball fans and France. These intuitions play a large role in many domains. 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.⁷ Those who use these methods 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 and elementary logic, they could demonstrate that the intuitions led to unambiguous errors. Unfortunately, that cannot be demonstrated here. In the moral and political domains, 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. 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 suggestion is that moral heuristics exist and are indeed pervasive. We should not treat the resulting moral intuitions as fixed points for analysis, rather than as unreliable and at least potentially erroneous.

I. ORDINARY HEURISTICS IN ACTION

A. AVAILABILITY

The classic work on heuristics and biases deals not with moral questions, but with issues of fact.⁸ With respect to heuristics, the basic claim 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

7. See Bernard Williams, *A Critique of Utilitarianism*, in J.J.C. SMART & BERNARD WILLIAMS, *UTILITARIANISM: FOR AND AGAINST* 149–50 (1973).

8. The key papers can be found in *JUDGMENT UNDER UNCERTAINTY*, *supra* note 3; a more recent collection is *HEURISTICS AND BIASES*, *supra* note 4. 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 & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, in *CHOICES, VALUES, AND FRAMES* 17–44 (2000).

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 letter⁹—even though a moment’s reflection shows that this cannot possibly be true. People err because they use an identifiable 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.¹⁰ 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.”¹¹ 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.

B. REPRESENTATIVENESS

The most famous example of the representativeness heuristic 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.”¹² People were asked to rank, in order of probability, eight possible futures for Linda.¹³ 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.¹⁴ 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

9. HEURISTICS AND BIASES, *supra* note 4, at 21.

10. See Tversky & Kahneman, *supra* note 3, at 3, 11–14.

11. *Id.* at 11.

12. Daniel Kahneman & Shane Frederick, *Representativeness Revisited: Attribute Substitution in Intuitive Judgment*, in HEURISTICS AND BIASES, *supra* note 4, at 62 (citing Tversky & Kahneman, *supra* note 3, at 92).

13. *Id.*

14. *Id.*

from the representativeness heuristic: Linda's description seems to match "bank teller and active in the feminist movement" far better than "bank teller."¹⁵

In an illuminating and even profound reflection on the example, Stephen Jay Gould observed, "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.'"¹⁶ We shall return to Gould's homunculus on several occasions.

The representativeness heuristic appears to explain many bad decisions by baseball scouts, who have in mind a "prototype" of a successful baseball player, and rely on the prototype to make decisions.¹⁷ Undoubtedly personnel decisions in many domains are based on prototypes in a way that produce many wise choices but also some unwise ones. The representativeness heuristic also helps explain what Paul Rozin and Carol Nemeroff call "sympathetic magical thinking," including the beliefs that some objects have contagious properties and that causes resemble their effects.¹⁸ Many educated Americans will not eat food touched by a sterilized cockroach.¹⁹ They refuse chocolates that have been shaped into realistic-looking dog feces.²⁰ They are reluctant to use sugar from a bottle labeled "Sodium Cyanide, Poison," even if they are assured, and believe, that the bottle really contains sugar and never contained cyanide—and indeed even if they themselves placed the label, arbitrarily, on that particular bottle!²¹ In fact people are reluctant to eat sugar labeled "Not Sodium Cyanide," apparently because the very words "Sodium Cyanide" automatically conjure negative associations.²²

I speculate that the immense popularity of organic foods

15. *Id.*

16. STEPHEN JAY GOULD, *BULLY FOR BRONTOSAURUS: REFLECTIONS IN NATURAL HISTORY* 469 (1991).

17. See MICHAEL LEWIS, *MONEYBALL: THE ART OF WINNING AN UNFAIR GAME* 15–20 (2003); see also Richard H. Thaler & Cass R. Sunstein, *Who's on First*, *THE NEW REPUBLIC*, Sept. 1, 2003, at 28.

18. See Paul Rozin & Carol Nemeroff, *Sympathetic Magical Thinking: The Contagion and Similarity "Heuristics,"* in *HEURISTICS AND BIASES*, *supra* note 4, at 201.

19. *Id.* at 202.

20. *Id.*

21. *Id.*

22. *Id.*

owes a great deal to the representativeness heuristic, above all because of the view that there is an association between the natural and the healthy, and between chemical and danger.²³ I speculate too that many unjustified judgments about security risks are fueled by the representativeness heuristic. The internment of Americans of Japanese descent, at the early stages of World War II, was undoubtedly connected with the use of that heuristic after the attack on Pearl Harbor. And might not the same heuristic help explain the widespread belief, held by a majority of Americans, that Saddam Hussein was involved in the attacks of September 11?²⁴

C. AFFECT

A great deal of recent attention has been paid to the fact that people often have a rapid, largely affective response to objects and situations, including job applicants, consumer products, animals, cars, and causes of action.²⁵ A jury might have an immediate negative reaction to a plaintiff in a personal injury case; a judge might have a positive intuitive reaction to an equal protection claim; an employer might instantly like, or dislike, someone who has applied for a job.²⁶

Can affect operate as a “heuristic”? Paul Slovic and his co-authors suggest an affirmative answer.²⁷ They contend that our affective responses occur rapidly and automatically, and that people use their feelings as a kind of substitute for a more systematic, all-things-considered judgment.²⁸ Unfortunately, there is an obvious sense in which it is unhelpful to treat “affect” as an explanation for someone’s attitude toward objects. In some settings, affect *represents*, or *is*, that very attitude, and there-

23. For criticism of that association, see JAMES COLLMAN, *NATURALLY DANGEROUS: SURPRISING FACTS ABOUT FOOD, HEALTH, AND THE ENVIRONMENT* 1–2 (2001) (challenging the belief that organic products are necessarily safe and wholesome).

24. Dana Milbank & Claudia Deane, *Hussein Link to 9/11 Lingers in Many Minds*, WASH. POST, Sept. 6, 2003, at A1 (noting that sixty-nine percent of Americans polled by the Washington Post believed that Saddam Hussein was linked to the September 11, 2001, terrorist attacks).

25. Paul Slovic et al., *The Affect Heuristic*, in *HEURISTICS AND BIASES*, *supra* note 4, at 414–15 (discussing the affect heuristic).

26. See Timothy Wilson et al., *Mental Contamination and the Debiasing Problem*, in *HEURISTICS AND BIASES*, *supra* note 4, at 185, 198–99.

27. Slovic et al., *supra* note 25, at 397.

28. *Id.*

fore cannot explain or account for it. (Would it be helpful to explain Tom's romantic attraction to Anne by saying that Anne produces a favorable affect in Tom?) The idea must be that affect works in the same way as availability and representativeness: in many contexts, people's emotional reactions are substituting for a more careful and deliberative inquiry into the (factual?) issues at stake.

The simplest way to establish this would be to proceed as Kahneman and Tversky originally did, by showing, for example, that people assess questions of probability by reference to affect, and that this method leads to predictable errors. What is the probability of death from smoking, driving, flying, or eating pesticides? If people's affect toward these activities matched their probability judgments, producing systematic error, it would certainly be plausible to speak of an affect heuristic. While we do not have data of this sort, there is some closely related evidence suggestive of an affect heuristic in the domain of risk. When asked to assess the risks and benefits associated with certain items, people tend to say that risky activities contain low benefits, and that beneficial activities contain low risks.²⁹ It is rare that they will see an activity as *both* highly beneficial and quite dangerous, or as both benefit-free and danger-free. Because risk and benefit are distinct concepts, this finding seems to suggest that "affect" comes first, and helps to "direct" judgments of both risk and benefit.

The claim is fortified by a study asking people to make decisions under time pressure.³⁰ The motivating claim is that the affect heuristic is more efficient than analytic processing in the sense that it permits especially rapid assessments. Under time pressure, the experimenters hypothesize that there would be an unusually strong inverse correlation between judged risk and judged benefit, because affect will be the determinant of assessment, and people will have less time to undertake the kind of analysis that could begin to pull the two apart.³¹ The hypothesis is confirmed: under time pressure, the inverse correlation is even stronger than without time pressure.³²

The affect heuristic raises many puzzles, because affect it-

29. *Id.* at 410–13.

30. *Id.* at 412.

31. *Id.*

32. *Id.* at 412–13.

self remains to be explained; undoubtedly the availability and representativeness heuristics play a role in producing affect. But an understanding of the role of affect casts a number of facts in a new light. Background mood, for example, influences decisions and reactions in many domains.³³ Consider the remarkable fact that stock prices increase significantly on sunny days, a fact that is hard to explain in terms that do not rely on affect.³⁴ Another study suggests that when people are anxious and fearful, they are less likely to engage in systematic processing and more prone to error.³⁵ Still another study finds that when people are afraid, they are more likely to engage in negative stereotyping of members of other social groups.³⁶

D. HEURISTICS AND DUAL PROCESSING

Kahneman and Shane Frederick have suggested that the heuristics should be seen in light of dual-process theories of cognition.³⁷ Those theories distinguish between two families of cognitive operations, sometimes labeled System I and System II. System I is intuitive; it is rapid, automatic, and effortless (and it features Gould's homunculus). System II, by contrast, is reflective; it is slower, self-aware, calculative, and deductive. Kahneman and Frederick suggest that System I proposes quick answers to problems of judgment, while System II operates as a monitor, confirming or overriding those judgments. Consider, for example, someone who is flying from Chicago to New York in the month after an airplane crash. This person might make a rapid, barely conscious judgment, rooted in System I, that the flight is quite risky; but there might well be a System II over-

33. See Alice M. Isen, *Positive Affect and Decision Making*, in RESEARCH ON JUDGMENT AND DECISION MAKING 509, 513 (William M. Goldstein & Robin Hogarth eds., 1997).

34. See David Hirschleifer & Tyler Shumway, *Good Day Sunshine: Stock Returns and the Weather* (2001), at <http://papers.ssrn.com/sol3/results.cfm>.

35. See Gordon B. Moskowitz et. al., *The History of Dual Process Notions: and the Future of Preconscious Control*, in DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY 12, 19–20 (Shelly Chaiken & Yaacov Trope eds., 1999) (noting that “people who are anxious about or vulnerable to a health threat, or otherwise experiencing stress may engage in less careful or less extensive processing of health relevant information”).

36. See William von Hippel et al., *Attitudinal Process Versus Context: The Role of Information Processing Biases in Social Judgment and Behavior*, in SOCIAL JUDGMENTS 251, 263 (Joseph P. Forgas et al. eds., 2003).

37. For overviews, see SOCIAL JUDGMENTS, *supra* note 36; DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY, *supra* note 35.

ride, bringing a more realistic assessment to bear.

Kahneman and Frederick also offer a general claim about the nature of heuristics: they operate through a process of *attribute substitution*.³⁸ In this process, people are interested in assessing a “target attribute,” and they do so by substituting a “heuristic attribute” of the object, which is easier to handle.³⁹ Consider the question whether more people die from suicides or homicides. Lacking statistical information, people might respond by asking whether it is easier to recall cases in either class (the availability heuristic). The response might well be sensible, but it might also lead to errors.

Is it possible that System I and System II operate in the moral domain as well? Might people have rapid, intuitive moral assessments, subject to correction by the more deliberative processes reflected in System II? Might moral judgments substitute a heuristic attribute for a target attribute? Might the same be true in legal contexts? There is every reason to suggest an affirmative answer to these questions. Juries are likely to be affected by simple heuristics about guilt or innocence and about the police; ordinary moral assessments, about who has done right and who has done wrong, are much influenced by the availability, representativeness, and affect heuristics.

II. MORAL HEURISTICS: A CATALOGUE

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.

A. EXCESSIVELY AMBITIOUS STARTS

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.⁴⁰ 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. Sup-

38. Kahneman & Frederick, *supra* note 12, at 51–52.

39. *Id.* at 53.

40. See BARON, JUDGMENT MISGUIDED, *supra* note 1, at 14–18; see also LOUIS KAPLOW & STEVEN SHAPELL, FAIRNESS VERSUS WELFARE (2002).

pose people are committed to retributivism as their preferred theory of punishment. Might they be making a cognitive error? (Is Kantianism a series of cognitive errors?) With respect to moral principles, deontologists and utilitarians typically agree about concrete cases; they can join in accepting the basic rules of criminal and civil law. When deontologists and others depart from utilitarian principles, perhaps they are operating on the basis of heuristics that usually work well but that sometimes misfire.

But easy victories are unlikely here. In the case of many ordinary heuristics, based on availability and representativeness, people can be readily embarrassed once it is shown that their judgments cannot possibly map onto reality.⁴¹ If they are not embarrassed, an easily accessible objective standard can demonstrate an error of fact or logic. In the moral domain, this is harder to demonstrate. To say the least, people who reject utilitarianism are not easily embarrassed by a demonstration that their moral judgments can lead to reductions in utility. In the moral domain, it is more difficult to produce a widely shared standard by which to test the question of mistake.

Consider, for example, the controversy over the effort by Louis Kaplow and Steven Shavell to demonstrate that principles not based entirely on welfare sometimes produce policies that make everyone worse off.⁴² Kaplow and Shavell urge that if the legal system attempts to promote fairness, it will undermine people's actual welfare—possibly the welfare of most or even all individuals.⁴³

Many critics have refused to accept the Kaplow/Shavell argument on behalf of welfare even if it is assumed that this argument is essentially correct on its own terms.⁴⁴ Indeed, those who reject utilitarianism might easily turn the tables. They

41. Note that this point does not hold for all heuristics. See Deborah Frisch, *Reasons for Framing Effects*, 54 *ORG. BEHAV. & HUMAN DECISION PROCESSES* 399, 421–23 (1993) (demonstrating that many heuristics are quite sticky, even when identified as such).

42. See KAPLOW & SHAVELL, *supra* note 40, at 52–58.

43. *Id.*

44. See, e.g., Richard H. Fallon, Jr., *Should We All Be Welfare Economists?*, 101 *MICH. L. REV.* 979, 980 (2003) (commenting that, “if accepted, Kaplow and Shavell’s argument would support recognition only of rights that tend to be utility-maximizing”); Jules L. Coleman, *The Grounds of Welfare*, 112 *YALE L.J.* 1511 (2003) (reviewing LOUIS KAPLOW & STEVEN SHAVELL, *FAIRNESS VERSUS WELFARE* (2002)).

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. On this view, utilitarianism is itself a heuristic, one that usually works well but leads to systematic errors.⁴⁵ 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.⁴⁶

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 of 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.⁴⁷ If people think that some person named Linda is more likely to be a “a bank teller who is active in the feminist movement” than a “bank teller,” there is an evident problem.⁴⁸ In the cases that concern me here, factual blunders and simple logic do not provide such a simple test.

In some particular cases, however, we might be able to make some progress by investigating less abstract commitments and entertaining the hypothesis that for them, some widely accepted rules are heuristics. Consider, for example, the commitment to colorblindness, or the view that free speech is “an absolute.” For those who believe that discrimination against African-Americans is far worse than discrimination against whites, the idea of colorblindness might seem to be a heuristic: an overgeneralization of a sound moral principle that usually works well, or that worked well in the context for which it was designed, but that misfires when applied to affirmative action. The idea that free speech is “an absolute,” or can never

45. See generally *LOCAL JUSTICE IN AMERICA* (Jon Elster ed., 1995) (discussing the overlap between utilitarianism and deontological accounts).

46. See, e.g., SMART & WILLIAMS, *supra* note 7 (providing both an outline and critique of utilitarian principles).

47. See *supra* note 9 and accompanying text.

48. See *supra* notes 12–16 and accompanying text.

be regulated, is palpably wrong. But, many people believe that they are committed to it, and perhaps the commitment is a heuristic that serves desirable social functions.

In fact many people show absolute commitments to various propositions. In one study, strong majorities of people said that the following items were unacceptable no matter how great the benefits: electing a politician who has made racist comments, cutting all trees in an old-growth forest, sterilizing retarded women, building a new water system with pipes found to contain lead, and medical removal of dying patients' organs without their consent.⁴⁹ But in the same study, people were also asked to think of counterexamples; when so asked, significant numbers of people changed their position, concluding that their commitment was not in fact absolute.⁵⁰ People were also sensitive to the probability that harm would occur. With respect to the protection of endangered species and ecological harms from genetic engineering, many people will say, in the abstract, that they will not tolerate tradeoffs.⁵¹ But if the probability of harm is low enough, people are willing to create risks to endangered species and of ecological harm.⁵² These findings strongly suggest that when people speak in absolute terms, they are generalizing from most situations, producing moral heuristics that predictably misfire.

Or turn to the question whether government is ever permitted to torture known (or suspected) terrorists, even in contexts in which torture is the only way to prevent a massive number of deaths. A flat prohibition on torture might be justifiable on ethical grounds. But perhaps any flat prohibition is really a heuristic, one that works well most or almost all of the time, but that also leads to mistakes in imaginable cases. Should government ever be permitted to experiment on human beings? What about severely disabled human beings with cognitive capacities below those of, say, chimpanzees? A ban on certain medical experiments on human beings might be a heuristic, unsuited to imaginable cases.

I strongly suspect that in all of these cases, heuristics are at work. Unfortunately, however, it is difficult to establish this

49. Jonathan Baron & S. Leshner, *How Serious Are Expressions of Protected Values?*, 6 J. EXPERIMENTAL PSYCHOL.: APPLIED 183, 185 (2000).

50. *Id.*

51. *Id.* at 186.

52. *Id.*

point without making highly contentious claims about the merits.⁵³ Some people believe that colorblindness is morally and legally required; they do not believe that they are using a heuristic. No one really thinks that free speech is an absolute; the implausibility of that belief is simple to demonstrate, and after thinking about the regulation of bribery, perjury, and conspiracy, Gould's homunculus is nowhere to be found. Those who oppose torture, and who object to any medical experimentation on human beings, would not be easily embarrassed by the imaginable cases concocted by their adversaries.

Let us consider, then, some more promising possibilities.⁵⁴ In most of the examples that follow, I believe that it is possible to conclude that a moral heuristic is at work without accepting any especially controversial normative claims. To accept some of these examples, it will be necessary for readers to accept no contestable substantive position at all. To accept others, it will be necessary to accept only one substantive position, which we might call "weak consequentialism." According to weak consequentialism as I am understanding it here, the social consequences of the legal system are relevant, other things being equal or nearly so, to what law ought to be doing. Weak consequentialists can reject utilitarianism; they might agree that a violation of rights is part of the set of consequences that must be considered. They argue only that the full set of consequences is relevant to what must be done. Weak consequentialism, in the form that I am understanding it here, seems to me sufficiently nonsectarian to support the idea that in the cases at hand, moral heuristics are playing a significant role.

B. POINTLESS PUNISHMENT AND INCENTIVE NEGLECT

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 wrong-

53. For strong evidence without such claims, see *id.* at 192 (asserting that people "will not accept actions that violate [protected values] if the probability or amount of the harm is small relative to the probability and magnitude of benefit").

54. On the question of whether moral commitments are as absolute as they appear, see Baron & Leshner, *supra* note 49, at 183–84, which finds two effective ways of challenging such commitments: the first is to ask people to think of counterexamples; the second is to give people extreme examples such as preventing individuals from shouting "fire" in a crowded theater.

doing. People are likely to want to punish the corporation as if it were a person.⁵⁵ 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*.⁵⁶ According to this heuristic, people’s punishment judgments are a product of their outrage.⁵⁷ This heuristic produces reasonable 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.⁵⁸ In one case, subjects were told that the result of a higher penalty would be to make the company try harder to make safer products.⁵⁹ In an adjacent case, subjects were told that a higher penalty would make the company more likely to stop making the product, with the result that fewer safe products would be on the market.⁶⁰ Most subjects, including a group of judges, gave the same penalties in both cases.⁶¹ “Most of the respondents did not seem to notice the incentive issue.”⁶² Let us call the resulting set of judgments a form of *incentive neglect*, which is apparently pervasive. In another study, people said that they would give the same punishment to a company that would respond with safer products and one that would be unaffected because the penalty would be

55. For evidence to this effect, see generally PUNITIVE DAMAGES: HOW JURIES DECIDE (Cass R. Sunstein et al. eds., 2002); Daniel Kahneman et al., *Shared Outrage and Erratic Awards: The Psychology of Punitive Damages*, 16 J. RISK & UNCERTAINTY 49, 58–72 (1998).

56. See Daniel Kahneman & Shane Frederick, *Representativeness Revisited: Attribute Substitution in Intuitive Judgment*, in HEURISTICS AND BIASES, *supra* note 4, at 49, 63–65.

57. *Id.*

58. Jonathan Baron & Ilana Ritov, *Intuitions About Penalties and Compensation in the Context of Tort Law*, 7 J. RISK & UNCERTAINTY 17 (1993).

59. *Id.* at 22–23.

60. *Id.*

61. *Id.* at 23–26.

62. See Jonathan Baron, *Heuristics and Biases in Equity Judgments: A Utilitarian Approach*, in PSYCHOLOGICAL PERSPECTIVES ON JUSTICE, *supra* note 6, at 109, 123 (discussing the study reported in Baron & Ritov, *supra* note 58).

secret (with those who would know retiring) and covered by insurance (whose price would not increase).⁶³ Here too the effects of the punishment did not affect judgments by a majority of respondents.⁶⁴

A similar result emerged from a test of punishment judgments that asked subjects, including judges and legislators, to choose penalties for dumping hazardous waste.⁶⁵ In one case, the penalty would make companies try harder to avoid waste.⁶⁶ In another, the penalty would lead companies to cease making a beneficial product.⁶⁷ Most people did not penalize companies differently in the two cases.⁶⁸ 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.⁶⁹ 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?

In these cases, I think that it is most sensible to think that people are operating under a heuristic, requiring punishment that is proportional to outrageousness, and thinking that punishment 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 regardless of whether the consequence would be to make human beings safer and healthier. Those who insist on proportional punishments might disagree; perhaps they would contend that in neglecting consequences, they are embracing a large principle to which they are committed. But it might be worthwhile for them to consider the possibility that they have been tricked by a heuristic—and that their reluctance to acknowledge the point might be a product not of a sound principle, but of the insistent voice of their own version of Gould's homunculus.

63. *Id.* at 123–24.

64. *Id.*

65. Jonathan Baron et al., *Attitudes Toward Managing Hazardous Waste: What Should Be Cleaned Up and Who Should Pay for It?*, 13 RISK ANALYSIS 183, 184 (1993).

66. *Id.* at 185.

67. *Id.*

68. *Id.* at 186–87.

69. *Id.* at 190–91.

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.⁷⁰ 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: when trust is betrayed, the damage is worse than when an otherwise identical act had been committed by someone who was not a beneficiary of trust. It should therefore not be surprising that people will favor greater punishment for betrayals than for otherwise identical crimes.⁷¹

The disparity can be plausibly justified on the ground that the 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? Along the same line, it is bad to have an unfaithful spouse, but it is even worse if the infidelity occurred with your best friend, because that kind of infidelity makes it harder to have trusting friendships in the future.

In this light, it is possible to understand why betrayals produce special moral opprobrium and, where the law has been violated, increased punishment. Consider, however, a finding that is much harder to explain: *People are especially averse to risks of death that come from products (like air bags) 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 air bag.⁷² The study involved

70. See generally Jonathan J. Koehler & Andrew D. Gershoff, *Betrayal Aversion: When Agents of Protection Become Agents of Harm*, 90 ORGANIZATIONAL BEHAV. & HUMAN DECISION PROCESSES 244 (2003) (relying on five empirical studies to find that acts of betrayal elicited stronger desired punishments than other bad acts).

71. *Id.* at 245. Similarly, a study of “aversion to safety product” betrayal risks found that “when faced with a choice among pairs of safety products . . . most people preferred inferior options (in terms of overall risk exposure) over those that were associated with a slim chance of betrayal.” *Id.* at 257.

72. *Id.* at 252–56.

two principal conditions. In the first, people were asked to choose between two equally priced cars, Car A and Car B.⁷³ According to crash tests, there was a 2% chance that drivers of Car A, with Air Bag A, will die in serious accidents as a result of the impact of the crash.⁷⁴ With Car B, and Air Bag B, there was a 1% chance of death, but also an additional 1 in 10,000 (0.01%) chance of death as a result of deployment of the air bag.⁷⁵ Similar studies involved vaccines and smoke alarms.⁷⁶

The result was that over two-thirds of participants chose the higher risk safety option when the less risky one carried a "betrayal risk."⁷⁷ A control condition demonstrated that people were not confused about the numbers: when asked to choose between a 2% risk and a 1.01% risk, people selected the 1.01% risk so long as betrayal was not involved.⁷⁸ 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. "Apparently, people are willing to incur greater risks of the very harm they seek protection from to avoid the mere possibility of betrayal."⁷⁹ Remarkably, "betrayal risks appear to be so psychologically intolerable that people are willing to double their risk of death from automobile crashes, fires, and diseases to avoid incurring a small possibility of death by safety device betrayal."⁸⁰

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 badly in some cases, as when those who deploy it end up increasing the risks they themselves face. An air bag 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 air bag makes people safer on balance, people should use it, even if in a tiny percent-

73. *Id.* at 254.

74. *Id.*

75. *Id.*

76. *Id.* at 254–55 (finding results similar to those using the air bag scenario).

77. *Id.* at 255.

78. *Id.* at 256.

79. *Id.* at 244.

80. *Id.* at 255.

age of cases it creates a risk that would not otherwise exist. To reject air bags 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 involving products and an increase in likelihood of death, is the moral cousin to the use of the representativeness heuristic in the Linda case.⁸¹

In a sense, the special antipathy to betrayal risks might be seen to involve not a moral heuristic but a taste. In choosing products, people are not making pure moral judgments; they are choosing what they like best. A purer test of a moral heuristic would involve an experiment asking whether people are averse to betrayal risks when they are purchasing safety devices for their friends or family members. I am unaware of no such direct test, but it is certainly reasonable to expect that it would not produce substantially different results from those in the experiments just described. In fact, closely related experiments support that expectation.⁸² In deciding whether to vaccinate their children from the risks of serious diseases, people report a form of "omission bias." Many people are more alert to the risks of the vaccination than to the risk from the disease, so much so that they will expose their children to a greater risk from "nature" than from the vaccine.⁸³ The omission bias reflects a moral heuristic, to the effect that *people ought not to take affirmative steps that will inflict risks on their own children*. The omission bias is closely related to the special antipathy to betrayals.

D. HUMAN CLONING AND TAMPERING WITH NATURE

By a large margin, most Americans reject human cloning and believe that it should be banned.⁸⁴ To say the least, the

81. See *supra* notes 12–15 and accompanying text.

82. See Ilana Ritov & Jonathan Baron, *Reluctance to Vaccinate: Omission Bias and Ambiguity*, in BEHAVIORAL LAW AND ECONOMICS 168, 184 (Cass R. Sunstein ed., 2002) (finding that "[s]ubjects are reluctant to vaccinate when the vaccine can cause bad outcomes, even if the outcomes of not vaccinating are worse").

83. *Id.*

84. According to a comparison of survey results compiled by the Center for Genetics and Society, when asked, "Should scientists be allowed to clone humans?," 85% of Americans in a May 2002 CBS News poll answered "no." In a Gallup poll for May 2002, 90% of Americans opposed cloning that is designed

ethical and legal issues here are extremely complex. It is necessary, for example, to distinguish between reproductive and nonreproductive cloning; the former is designed to produce children, whereas the latter is designed to produce cells for therapeutic use. With respect to reproductive cloning, it is important to identify the particular grounds for moral concern. Do we fear that cloned children would be means to their parents' ends, and if so why? Do we fear that they would suffer particular psychological harm? Do we fear that they would suffer from especially severe physical or psychological problems?

In a discussion of the problem of new reproductive technologies, above all cloning, Leon Kass points to the "wisdom of repugnance."⁸⁵ Kass writes:

People are repelled by many aspects of human cloning. They recoil from the prospect of mass production of human beings, with large clones of look-alikes, compromised in their individuality; the idea of father-son or mother-daughter twins; the bizarre prospects of a woman's giving birth to and rearing a genetic copy of herself, her spouse, or even her deceased father or mother; the grotesqueness of conceiving a child as an exact replacement for another who has died; the utilitarian creation of embryonic genetic duplicates of oneself, to be frozen away or created when necessary, in case of need for homologous tissues or organs for transplantation; the narcissism of those who would clone themselves and the arrogance of others who think they know who deserves to be cloned or which genotype any child-to-be should be thrilled to receive; the Frankensteinian hubris to create human life and increasingly to control its destiny; man playing God. . . .

We are repelled by the prospect of cloning human beings not because of the strangeness or novelty of the undertaking, but because we intuit and feel, immediately and without argument, the violation of things that we rightfully hold dear. . . . Shallow are the souls that have forgotten how to shudder.⁸⁶

Kass is correct to suggest that revulsion toward human cloning might be grounded in legitimate concerns, and I mean to be agnostic here on the question whether human cloning is

to specifically result in the birth of a human being, and 61% opposed cloning of human embryos for use in medical research. According to a February 2001 Time/CNN poll, 92% of surveyed Americans were of the opinion that the prospect of creating genetically superior human beings did not justify creating a human clone. CENTER FOR GENETICS AND SOCIETY, Summary of Survey Results, at <http://www.genetics-and-society.org/analysis/opinion/summary.html> (last visited Mar. 6, 2004).

85. Leon R. Kass, *The Wisdom of Repugnance*, in LEON R. KASS & JAMES Q. WILSON, *THE ETHICS OF CLONING* 3, 17–24 (1998).

86. *Id.* at 17–19.

ethically defensible. But I want to suggest that moral heuristics might well be responsible, for Kass seeks to celebrate as “we intuit and feel, immediately and without argument.”⁸⁷ In fact, Kass’s catalogue of alleged errors seems to me an extraordinary exercise in the use of such heuristics. Availability operates in this context, not to drive judgments about probability, but to call up instances of morally dubious behavior (for example, “mass production of human beings, with large clones of look-alikes, compromised in their individuality”).⁸⁸ The representativeness heuristic plays a similar role (for example, “the Frankensteinian hubris to create human life and increasingly to control its destiny”).⁸⁹ But I believe that Kass gets closest to the cognitive process here with three words: “man playing God.”⁹⁰

In many contexts, a simple heuristic plays a large role in judgments of fact and morality: *Do not tamper with nature*. That heuristic helps explain many risk-related judgments. For example, “[h]uman intervention seems to be an amplifier in judgments on food riskiness and contagion,” even though “more lives are lost to natural than to man-made disasters in the world.”⁹¹ Studies show that people overestimate the carcinogenic risk from pesticides and underestimate the risks of natural carcinogens.⁹² They also believe that nature implies safety, so much that they will prefer natural water to processed water even if they are chemically identical.⁹³ The moral injunction against tampering with nature plays a large role in public objections to genetic engineering of food, and hence legal regulation of such engineering is sometimes driven by that heuristic, rather than by a deliberative encounter with the substantive issues.⁹⁴ The idea that we should not tamper with nature is so

87. *Id.* at 19.

88. *Id.* at 17.

89. *Id.* at 18.

90. *Id.*

91. Paul Rozin, *Technological Stigma: Some Perspectives from the Study of Contagion*, in *RISK, MEDIA, AND STIGMA* 31, 38 (James Flynn et al. eds., 2001).

92. *Id.*

93. *Id.*

94. See ALAN MCHUGHEN, *PANDORA’S PICNIC BASKET: THE POTENTIAL AND HAZARDS OF GENETICALLY MODIFIED FOODS* 230–42 (2000) (providing evidence that organic foods are not necessarily safer than genetically modified foods).

wildly implausible that it cannot account for the intensity of public opposition to cloning. There appears to be a narrower heuristic at work, bred in part, perhaps, from the cognitively available tale of Frankenstein's monster: *Do not tamper with natural processes for human reproduction*. It is not clear that this heuristic works well, but it is clear that it systematically misfires.

E. 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 life saved—a ceiling that is, by the way, significantly higher than the amount the Environmental Protection Agency uses for a statistical life.⁹⁵ How will ordinary people react to this decision? The answer is that they will not react favorably.⁹⁶ In fact, people tend to punish companies that base their decisions on cost-benefit analysis, even if a high valuation is placed on human life. By contrast, they impose less severe punishment on companies that are willing to impose a "risk" on people but that do not produce a risk analysis that measures lives lost and dollars, and trades one against another.⁹⁷ What underlies these moral judgments?

A careful look raises the possibility that when people disapprove of trading money for lives, 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*. In ordinary life, you should not engage in conduct with the knowledge that several people will die as a result. If you are playing some sport or are working on your yard, you ought not to continue if you believe that your actions will kill others. Invoking

95. See Cass R. Sunstein, *The Arithmetic of Arsenic*, 90 GEO. L.J. 2255, 2274 (2002) (citing the EPA amount of \$6.1 million per life saved).

96. See W. Kip Viscusi, *Corporate Risk Analysis: A Reckless Act?*, 52 STAN. L. REV. 547, 556–58 (2000) (observing that jury awards tend to be higher when companies base decisions on cost-per-life-saved assessments).

97. See *id.*; see also Philip Tetlock, *Coping With Tradeoffs*, in *ELEMENTS OF REASON: COGNITION, CHOICE, AND THE BOUNDS OF RATIONALITY* 239 (Arthur Lupia et al. eds., 2000).

that idea, people disapprove of companies that fail to improve safety when they are fully aware that deaths will result. By contrast, people do not disapprove of those who fail to improve safety while believing that there is a “risk” but 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.”⁹⁸ Companies that are aware that a “risk” exists but fail to do such analysis do not make clear, to themselves or 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.⁹⁹

Note that it is easy to reframe a probability as a certainty, and vice-versa; if I am correct, the reframing is likely to have large effects. Consider two cases:

- A. Company A knows that its product will kill ten people. It markets the product to its ten million customers with that knowledge. The cost of eliminating the risk would have been \$100 million.
- B. Company B knows that its product creates a one in one million risk of death. Its product is used by ten million people. The cost of eliminating the risk would have been \$100 million.

I have not collected data, but I am willing to predict, with a high degree of confidence, that Company A would be punished more severely than Company B, even though there is no difference between the two.

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. Of course this heuris-

98. It is also the case that explicit trading of money for lives is strongly disfavored. See Tetlock, *supra* note 97, at 239. 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.

99. Compare here the “identifiable victim effect”: people will expend a great deal to prevent a certain death to a single identifiable victim, and far less to eliminate a statistically larger harm in the form, for example, of a risk of 1/10,000 faced by 100,000 people. See generally Deborah A. Small & George Loewenstein, *Helping a Victim or Helping the Victim: Altruism and Identifiability*, 26 J. RISK & UNCERTAINTY 5 (2003); Karen E. Jenni & George Loewenstein, *Explaining the “Identifiable Victim Effect,”* 14 J. RISK & UNCERTAINTY 235 (1997).

tic does a great deal of good.¹⁰⁰ 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. Much of what is done by both industry and government is likely to result in one or more deaths. Of course it might well make sense in most or all of these domains to take extra steps to reduce risks. That proposition does not, however, support the implausible claim that we should disapprove 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 here too, "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 System I, and a crude but quite tenacious moral heuristic.

F. EMISSIONS TRADING

In the last decades, those involved in enacting and implementing environmental law have experimented with systems of "emissions trading."¹⁰¹ In those systems, polluters are typically given a license to pollute a certain amount, and the licenses can be traded on the market. The advantage of emissions trading systems is that if they work well, they will ensure emissions reductions at the lowest possible cost.

Is emissions trading immoral? Many people believe so.¹⁰² Michael Sandel, for example, urges that trading systems "un-

100. 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 infra* Part IV.

101. *See* CASS R. SUNSTEIN, *RISK AND REASON* 274-76 (2002).

102. *See* Michael Sandel, *It's Immoral to Buy the Right to Pollute*, N.Y. TIMES, Dec. 15, 1997, at A23; *see also* STEVEN KELMAN, WHAT PRICE INCENTIVES? 28-91 (1982) (rejecting an economic incentive approach to emission standards and outlining the purported social costs).

dermine the ethic we should be trying to foster on the environment."¹⁰³ He contends:

[T]urning pollution into a commodity to be bought and sold removes the moral stigma that is properly associated with it. If a company or a country is fined for spewing excessive pollutants into the air, the community conveys its judgment that the polluter has done something wrong. A fee, on the other hand, makes pollution just another cost of doing business, like wages, benefits and rent.¹⁰⁴

In the same vein, he objects to proposals to open carpool lanes to drivers without passengers who are willing to pay a fee. Here, as in the environmental context, it is undesirable to permit people to do something that is morally wrong so long as they are willing to pay for the privilege.

I believe that like other critics of emissions trading programs, Sandel is using a moral heuristic and that he has been fooled by his homunculus. The heuristic is this: *People should not be permitted to engage in moral wrongdoing for a fee.* You are not allowed to assault someone so long as you are willing to pay for the right to do so; there are no tradable licenses for rape, theft, or battery. The reason is that the appropriate level of these forms of wrongdoing is zero (putting to one side the fact that enforcement resources are limited; if they were unlimited, we would want to eliminate, not merely to reduce, these forms of illegality). But pollution is a different matter. At least some level of pollution is a byproduct of desirable social activities and products, including automobiles and power plants. Certain acts of pollution, including those that violate the law or are unconnected with desirable activities, are certainly wrong, but the same cannot be said of pollution as such. When Sandel objects to emissions trading, he is treating pollution as equivalent to a crime in a way that overgeneralizes a moral intuition that makes sense in other contexts. There is no moral problem with emissions trading as such. The insistent objection to emissions trading systems stems from a moral heuristic.

G. 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.¹⁰⁵ In one case, for

103. Sandel, *supra* note 102.

104. *Id.*

105. See, e.g., RONALD DWORKIN, FREEDOM'S LAW 130-46 (1996) (discuss-

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. Indeed, American constitutional law reflects judgments to this effect: people have a constitutional right to withdraw equipment that is necessary to keep them alive, but they have no constitutional right to physician-assisted suicide.¹⁰⁶ What exactly is the morally relevant difference?

It is worth considering the possibility that the act-omission distinction operates as a heuristic for the more complex and difficult assessment of the moral issues at stake.¹⁰⁷ 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 a bystander may do no such thing. (This is an overgeneralization, of course.) 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.¹⁰⁸ 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

ing the distinction between acts and omissions in the context of euthanasia).

106. See *Washington v. Glucksberg*, 521 U.S. 702, 724–25 (1997).

107. See BARON, *JUDGMENT MISGUIDED*, *supra* note 1, at 95–103 (1998) (discussing people's tendency to not feel responsible for harm caused by their failure to act).

108. See Baron, *Nonconsequentialist Decisions*, *supra* note 1, at 7–8.

persistent acceptance of withdrawal of life-saving equipment, alongside persistent doubts about euthanasia, is a demonstration of the point. There is no morally relevant difference between the two; the act-omission distinction makes a difference apparent or even clear when it is not real.¹⁰⁹

Consider in this regard the dispute over two well-known problems in moral philosophy. The problems do not quite involve the act-omission distinction, but they implicate closely related concerns. 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.¹¹⁰ 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.¹¹¹ 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.¹¹²

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. As a matter of principle, it is worse to throw a human being in the path of a trolley than to throw a switch that, perhaps indirectly, leads to a death. Hence a heuristic is at work: *Do not take active steps, by your own hands, to kill innocent people*. The relevant heuristic generally points in the right direction. To say the least, it is desirable for people to act on the basis of a moral heuristic that makes it extremely difficult to throw innocent people to their deaths. But the underlying heu-

109. There might, however, be practical reasons to believe that the right to physician-assisted suicide would be more subject to abuse than the right to withdrawal of life-saving equipment. See CASS R. SUNSTEIN, ONE CASE AT A TIME (1999).

110. See Judith Jarvis Thomson, RIGHTS, RESTITUTION AND RISK 39 (1986).

111. *Id.* at 82–83.

112. See, e.g., *id.* at 94–116.

ristic misfires 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 are not. In this sense, a moral heuristic leads to errors. It is good that people are extremely reluctant to throw people in the path of trains, but in imaginable cases, the extreme reluctance produces moral blunders. This objection does not bear only on ingeniously devised hypothetical cases. It suggests that a moral mistake pervades both common-sense morality and law, including constitutional law, by treating harmful omissions as morally unproblematic or categorically different from harmful actions.

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.¹¹³ 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,”¹¹⁴ and that brain areas associated with emotion are far more active in contemplating the footbridge problem than in contemplating the trolley problem.¹¹⁵ Of course this experiment is far from 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 consider the case of fear, where an identifiable region of the brain makes helpfully immediate but not entirely reliable judgments,¹¹⁶ in a way that suggests a possible physical location for some of the operations of System I. Immediate fear is often corrected by more reflective System II-like processes occupying other parts of the brain, in which people conclude, for example, that airplane travel is not dangerous after all, or that a loud noise is un-

113. See Joshua D. Greene et al., *An fMRI Investigation of Emotional Engagement in Moral Judgment*, 293 SCI. 2105 (2001).

114. *Id.* at 2107.

115. *Id.* at 2106.

116. See JOSEPH LEDOUX, *THE EMOTIONAL BRAIN* 166–74 (1998).

threatening. So too, perhaps, in the context of morality, politics, and law.

H. PROBABILITY OF DETECTION

Now turn to another example from the domain of punishment. On the economic account, the state's goal when imposing penalties for misconduct is to ensure optimal deterrence.¹¹⁷ 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 amount of the compensatory award. The simple exercise in multiplication will ensure optimal deterrence.

There remains a large question whether people accept this account, and if not, why not. (For the moment, let us put to one side the question whether they should accept it in principle.) Experiments suggest that people reject optimal deterrence and do not believe that the probability of detection is relevant to punishment. The reason is that they use the outrage heuristic.¹¹⁸ I participated in two experiments designed to cast light on this question.¹¹⁹ In the first, we gave people cases of wrongdoing, arguably calling for punitive damages, and also provided explicit information about the probability of detection.¹²⁰ The participants saw the same case, with only one difference: the probability of detection was substantially varied. People were asked about the amount of punitive damages that they would

117. A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 870–76 (1998).

118. See Kahneman & Frederick, *supra* note 56, at 63–65.

119. See Cass R. Sunstein et al., *Do People Want Optimal Deterrence?*, 29 J. LEGAL STUD. 237 (2000).

120. *Id.* at 241–42.

choose to award.¹²¹ 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.¹²² 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.¹²³ Even when their attention was explicitly directed to the probability of detection, people were indifferent to it. Their 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.¹²⁴ 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.¹²⁵

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 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 impose 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

121. *Id.*

122. *Id.* at 244–45.

123. *Id.* at 243.

124. *Id.* at 245–46.

125. *Id.* at 246.

with a company whose wrongdoing would inevitably be detected and punished.

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. Weak consequentialists do not reject retributivism, but they do claim that social welfare matters, and that it is appropriate to impose more punishment on people where the goal of deterrence requires it. 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. MORAL HEURISTICS, MARRIAGE, AND SEX

Issues at the intersection of morality, marriage, and sex provide an obvious place for the investigation of moral heuristics. These issues are of course sharply contested, but it is reasonable to hypothesize that at least some of the relevant moral judgments reflect the operation of moral heuristics. Such heuristics are peculiarly likely to be at work in any area in which people are likely to think, "That's disgusting!" Recall the earlier discussion of the question of cloning, where revulsion is a highly unreliable guide to moral judgment.¹²⁶

Any examples here will be contentious (and potentially embarrassing), but consider the incest taboo. We can easily imagine incestuous relationships, say between first cousins or second cousins, that ought not to give rise to social opprobrium, but that might nonetheless run afoul of social norms or even the law. The incest taboo is best defended by reference to coercion, psychological harm, and risks to children who might result from incestuous relationships. But in many imaginable cases, these concrete harms are not involved. Of course it is plausible to say that the best way to defend against these harms is by a flat prohibition on incest, one that has the disad-

126. See *supra* Part II.D.

vantage of excessive generality but the advantage of easy application. So defended, however, the taboo stands unmasked as a moral heuristic. In fact people seem to oppose incestuous relationships even if they are assured that no harms will come of them, and their opposition stays firm even though they are unable to justify it.¹²⁷ Many taboos, in this domain, can be analyzed in similar terms.¹²⁸

IV. EXOTIC CASES AND MORAL JUDGMENTS: A METHODOLOGICAL DIGRESSION

Some of these examples will seem more controversial than others. Taken as a whole, however, 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 rarely if ever encountered in ordinary life. Should you shoot an innocent person if that is the only way to save twenty innocent people?¹²⁹ 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?¹³⁰ Is there a difference between killing someone by throwing them into the path of a train and killing someone by diverting the train's path to send it in someone's direction?¹³¹

I believe that in cases of this kind, the underlying moral intuitions ordinarily work well, but 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 necessary to save others.* (I put to one side the contexts of self-defense and war.) In all likelihood, a society does much better if most people have this intuition, if only because judgments about necessity are likely

127. Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCH. REV. 814, 814–15 (2001).

128. See, e.g., Elizabeth Emens, *Monogamy's Law: Compulsory Monogamy and Polyamorous Existence*, 29 N.Y.U. REV. L. & SOC. CHANGE (forthcoming 2004).

129. Cf. Williams, *supra* note 7, at 97–100 (describing several examples of moral dilemmas potentially encountered in real life).

130. See THOMSON, *supra* note 110, at 31.

131. See *supra* notes 110–11 and accompanying text.

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 out 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 intuitions are reliable and should form the building blocks for sound moral judgments. An understanding of how heuristics work suggests reason to doubt the reliability of those intuitions, even when they are very firm.

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 is not.

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. Those who insist on an incest taboo might produce a better situation than those who urge that the taboo should be applied only in cases in which it is supported by its underlying rationale. 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 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 claim only that such heuristics lead to significant errors and a great deal of confusion.

How do these points bear on the search for reflective equilibrium? In John Rawls's influential formulation, people's judgments about justice should be made via an effort to ensure principled consistency between their beliefs at all levels of generality.¹³² All beliefs are revisable in principle. At the same time, some of our beliefs, about particular cases and more generally, seem especially fixed, and it will take a great deal to revise them. Nothing said here should be taken as a basis for questioning the search for reflective equilibrium. But if moral heuristics are pervasive, then some of our apparently fixed beliefs might be a product of them; we should be aware of that fact in attempting to reach reflective equilibrium. Judgments that seem to be most insistent, or least revisable, may result from overgeneralizing intuitions that work well in many contexts but that also misfire.

V. MORAL FRAMING

In cognitive psychology and behavioral economics, empirical work on heuristics has been accompanied by a literature on framing effects.¹³³ For a simple example, consider the question whether to undergo a risky medical procedure. When people are told, "Of those who have this procedure, ninety 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, ten percent are dead after five years."¹³⁴ Experience might be expected to solve this problem, but doctors too are vulnerable to this framing effect.¹³⁵ Here the question does not involve moral and political issues. Might framing effects be found there as well?

A. THE ASIAN DISEASE PROBLEM

Kahneman and Tversky themselves find moral framing effects in the context of what has become known as "the Asian

132. See JOHN RAWLS, A THEORY OF JUSTICE (1971).

133. See, e.g., Donald Redelmeier et al., *Understanding Patients' Decisions: Cognitive and Emotional Perspectives*, 270 JAMA 72, 73 (1993).

134. See *id.*

135. *Id.*

disease problem.”¹³⁶ Here is the first component of the problem:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Which of the two programs would you favor?

Most people choose Program A.¹³⁷

Now consider the second component of the problem, in which the same situation is given but followed by this description of the alternative programs:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

Most people choose Program D.¹³⁸ But a moment's reflection should be sufficient to show that Program A and Program C are identical, and so too for Program B and Program D. These are merely different descriptions of the same programs. The merely semantic shift in framing is sufficient to produce different outcomes. People's moral judgments about appropriate programs depend on whether the results are described in terms of "lives saved" or instead "lives lost." What accounts for the shift? The most sensible answer begins with the fact that people are risk-averse with respect to gains, but risk-seeking with respect to losses¹³⁹—what counts as a gain or a loss depends on the baseline from which measurements are made. Purely semantic framing can alter the baseline and, hence, alter moral judgments.

B. FUTURE GENERATIONS

A similar framing effect has been demonstrated in the im-

136. Kahneman & Tversky, *supra* note 8, at 4–5.

137. *Id.* at 5.

138. *Id.*

139. *See id.*

portant context of obligations to future generations,¹⁴⁰ a much-disputed question of morality, politics, and law.¹⁴¹ 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.¹⁴² In any case, an important question, asked in many debates about the issue, involves the nature of people's moral judgments on that issue. Indeed, an influential set of studies finds that people believe that it is morally appropriate to prefer risking few lives in the current generation to many lives of those in future generations.¹⁴³ From a series of surveys, Maureen Cropper and her coauthors suggest that people prefer saving 100 lives today to saving 7000 lives in 100 years.¹⁴⁴ They make this suggestion on the basis of answers to questions asking whether people would choose a program that saves "100 lives now" or a program that saves a substantially larger number "100 years from now."¹⁴⁵

It turns out, however, that the findings by Cropper and her coauthors are a product of a particular way of framing the problem, and that other ways of framing the same problem yield radically different results.¹⁴⁶ For example, most people consider "equally bad" a single death from pollution next year and a single death from pollution in 100 years.¹⁴⁷ This finding implies no preference for members of the current generation. In addition, people are equally divided between two programs: one that will save 55 lives now and 105 more lives in 20 years; and one that will save 100 lives now and 50 lives 25 years from now.¹⁴⁸ This finding also suggests no strong preference for

140. See Shane Frederick, *Measuring Intergenerational Time Preference: Are Future Lives Valued Less?*, 26 J. RISK & UNCERTAINTY 39, 48 (2003).

141. E.g., Richard Revesz, *Environmental Regulation, Cost-Benefit Analysis, and the Discounting of Human Lives*, 99 COL. L. REV. 941 (1999); Comment, *Judicial Review of Discount Rates Used in Regulatory Cost-Benefit Analysis*, 65 U. CHI. L. REV. 1333 (1998).

142. See Revesz, *supra* note 141, at 987-1016.

143. See Maureen Cropper et al., *Preferences for Life Saving Programs: How the Public Discounts Time and Age*, 8 J. RISK & UNCERTAINTY 243 (1994).

144. *Id.* at 251.

145. *Id.* at 246.

146. See Frederick, *supra* note 140.

147. *Id.* at 43.

148. *Id.* at 44.

members of the current generation.¹⁴⁹

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.¹⁵⁰ Consider the choice 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 prefer the first program, apparently suggesting that future lives are valued more highly.¹⁵¹ The reason for this seemingly odd result is that people like things to get better rather than worse, and they apply this preference, itself a kind of heuristic, to settings in which it makes little sense. In the same vein, people tend to prefer increasing wage profiles, even at their own economic expense. A five-year salary profile of \$90,000, \$80,000, \$70,000, \$60,000, and \$50,000 is generally rejected in favor of one of \$50,000, \$60,000, \$70,000, \$80,000, and \$90,000, even though the former is worth more.¹⁵² This preference, produced by a heuristic, is translated into the moral domain, thus making it possible to frame questions so that saving lives in the future is preferable to saving lives immediately.

The simplest conclusion is that people's moral judgments about obligations to future generations are very much a product of framing effects.¹⁵³ On this contested issue, there is no clear or robust judgment that can be elicited through moral questions. Answers are constructed, rather than elicited, in the process of response.¹⁵⁴

149. See *id.*

150. *Id.* at 45.

151. *Id.*

152. See George Loewenstein & Nachum Sicherman, *Do Workers Prefer Increasing Wage Profiles?*, 9 J. LAB. ECON. 67, 71-75 (1991); George Loewenstein & Drazen Prelec, *Negative Time Preference*, 81 AM. ECON. REV. 347, 347 (1991).

153. For a similar result, see Jonathan Baron, *Can We Use Human Judgments to Determine the Discount Rate?*, 20 RISK ANALYSIS 861, 866 (2000).

154. Compare John W. Payne et al., *Measuring Constructed Preferences: Towards a Building Code*, 19 J. RISK & UNCERTAINTY 243, 266 (1999), for an argument that in many domains, preferences are constructed rather than elicited by social situations. I am suggesting that the same is true for many moral issues as well.

C. LIVES AND LIFE-YEARS

Consider, as a final example, the question whether in valuing human life, government should take account of the number of "life-years" saved, or instead use a single number for every statistical life saved, regardless of the age of the beneficiaries. Some regulatory programs benefit people who are relatively young; others benefit people who are relatively old. If a program would prevent fifty deaths of people who are twenty, should it be treated the same way as a program that would prevent fifty deaths of people who are seventy? At least since 1976, analysts have suggested the possibility of focusing regulatory policy on either life-years or quality-adjusted life-years.¹⁵⁵ The issue received a great deal of public attention in connection with recent debates over the value of statistical life (VSL) and the value of a statistical life-year (VSLY) within the Office of Management and Budget (OMB) and the Environmental Protection Agency (EPA).¹⁵⁶ OMB has strongly encouraged federal agencies, including the EPA, to consider VSLY,¹⁵⁷ and OMB's draft guidelines on cost-benefit analysis ask agencies to "consider providing estimates of both VSL and VSLY."¹⁵⁸

Does OMB's view fit with or violate most people's moral judgments? The choice between statistical lives or VSL and statistical life-years or VSLY very much depends on framing. It should be easy to construct questions that would yield a preference for VSL:

Government is considering a policy that would count the value of elderly people as significantly less than the value of younger people. According to one proposal, for every dollar that people under seventy are worth, people over seventy are worth fifty-three cents. Do you approve of this proposal?

155. Richard Zeckhauser & Donald Shepard, *Where Now for Saving Lives?*, LAW & CONTEMP. PROBS., Autumn 1976, at 5.

156. For a discussion, see Cass R. Sunstein, *Lives, Life-Years, and Willingness to Pay*, COLUM. L. REV. (forthcoming 2004); Robert H. Hahn & Scott Wallsten, *Is Granny Worth \$2.3 Million or \$6.1 Million?*, at <http://www.aei.brookings.org/policy/page.php?id=138> (last visited Feb. 26, 2004).

157. See Hahn & Wallsten, *supra* note 156; Dana Wilkie, *White House Continues to Push for "Age" Discounts in Rulemaking*, COPLEYS NEWS SERV., May 16, 2003, available at <http://www.lexis.com/research>.

158. See Office of Mgmt. and Budget, Draft 2003 Report to Congress on the Costs and Benefits of Federal Regulations, 68 Fed. Reg. 5492, 5521 (Feb. 3, 2003).

I have not asked people to answer this question, but it is safe to predict that most respondents would answer "No." It should also be easy to construct questions that would suggest public disapproval of a uniform VSL:

Would you favor (a) a program that would save one hundred children from dying of a fatal cancer at the age of ten or instead (b) a program that would save one hundred and one senior citizens from dying of a fatal cancer at the age of eighty?

We can safely predict that most people would favor (a). In fact I have conducted a small survey myself, asking University of Chicago law students whether they would favor a policy that saves twenty people with a median age of forty or one that saves thirty people with a median age of sixty-five. By a majority of about two-to-one (fifty-three to twenty-five), students favored the former policy. As in the context of harms to future generations, highly variable responses should be expected in accordance with the nature of the question. It is extremely doubtful that people have stable, well-considered judgments on the issue.

D. MORAL FRAMING: WHAT FOLLOWS?

In one respect, the effects of framing should be entirely unsurprising. In the domain of polling and surveys, it is well known that responses will depend on the way the question is framed.¹⁵⁹ What I am emphasizing here is that in many areas, people's moral assessments are labile; they are constructed rather than elicited by the question. What follows from this claim?

The most obvious implications are prescriptive: it is not difficult to pose questions in a way that will move judgments in the preferred direction. A key factor here is *loss aversion*. People do not like losses from the status quo, and in fact they dislike losses more than they like corresponding gains.¹⁶⁰ Moreover, circumstances can be framed so that changes appear to be losses rather than gains. Consider two possible ways of encouraging people to use energy-conservation techniques:

159. For overviews, see ROGER TOURANGEAU ET AL., *THE PSYCHOLOGY OF SURVEY RESPONSE* (2000); Marianne Bertrand & Sendhil Mullainathan, *Do People Mean What They Say? Implications for Subjective Survey Data* (Jan. 2002) (unpublished working paper), http://papers.ssrn.com/paper.taf?abstract_id=260131.

160. See RICHARD H. THALER, *QUASI RATIONAL ECONOMICS* 5–10 (1991).

- A. If you use energy-conservation techniques, you will save \$X each year.
- B. If you do not use energy-conservation techniques, you will lose \$X per year.

It turns out that the second frame, using loss aversion, produces a larger behavioral shift than the first.¹⁶¹ The same point holds in the legal context. Suppose that a judge or jury is asked to assign compensatory damages for an injury in a tort case. Is the question: *How much money is needed to make the plaintiff as well off as he was before the accident occurred?* Or is it instead: *How much money would the plaintiff have to have been offered to accept the injury willingly?* According to conventional theory, the answer to the two questions should be the same; in fact this is the same question, differently framed. Nevertheless, empirical work shows that the second question will produce significantly higher figures.¹⁶²

In short, it is easy to frame questions in order to shift moral and legal assessments in the desired directions. But the normative implications of moral framing are unclear. It might be that in some domains, people lack robust or well-considered moral judgments, and what they conclude is inevitably a product of framing effects. If so, those who seek to elicit people's judgment might well use multiple frames, so as to reduce potential distorting effects.¹⁶³ Recall that both patients and doctors are more likely to choose a medical procedure if told that ninety percent of patients are alive after five years than if told that ten percent of patients are dead after that period.¹⁶⁴ Apparently people focus on the successes or the failures if they are informed about either. The identification of one or the other conjures up an immediate image of life or death, and that image plays a role in ultimate judgments.¹⁶⁵

To eliminate the distorting effect of the immediate image—a kind of System I distortion—it is appropriate to expose people

161. Cf. ELLIOT ARONSON, *THE SOCIAL ANIMAL* 124–25 (7th ed. 1995) (describing the analogous concept of contrast effects).

162. See Edward J. McCaffery et al., *Framing the Jury: Cognitive Perspective on Pain and Suffering Awards*, in *BEHAVIORAL LAW AND ECONOMICS* 259, 276 (Cass R. Sunstein ed., 2000).

163. Some lessons emerge from Payne et al., *supra* note 154, at 249–60.

164. See *supra* note 134 and accompanying text.

165. Hence, the result of this framing exercise seems to me connected to the findings discussed in Cass R. Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, 112 *YALE L.J.* 61, 74–83 (2002).

to more than one frame. In the context of the interests of future generations and the problem of lives versus life-years, some such approach is clearly desirable. Once it is used, the question is not a simple one of eliciting people's preferences. It is instead the beginning of a more complex process of deliberation and reflection.

CONCLUSION

My principal argument in this essay has been that moral heuristics and moral framing play a large role in social judgments. Sometimes this is not a problem. In many settings, people are subject to multiple framing effects, and it is possible to think in a way that reduces the distortion introduced by any single frame. Moral heuristics usually work well, and they can make it easy to arrive at moral judgments without requiring a complex and time-consuming inquiry into particular situations. Even better, moral heuristics can operate in a way that reduces the effects of venality, self-interest, or other biases that would undoubtedly confound moral assessments of particular cases. Along the same lines, moral heuristics have some of the virtues of rules. They outrun their rationale, and they are overinclusive; but they might work far better than an approach that tries to investigate, in individual circumstances, whether their rationale actually applies.

The problem is that moral heuristics, like their factual cousins, can produce large and systematic errors. If this 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. With respect to large-scale disputes about what morality requires, it is too contentious to argue that one side is operating under a heuristic, whereas another side has it basically right. But we should avoid undue skepticism here. Frequently moral evaluations can be made on grounds that seem, or should be, acceptable to all or almost all, and if so, we have a point of entry for the investigation of moral heuristics. I hope that I have said enough to show that in particular cases, sensible rules of thumb lead to demonstrable errors not merely in factual judgments, but in the domains of morality, politics, and law as well.