Is Deontology a Heuristic? On Psychology, Neuroscience, Ethics, and Law

Citation

Cass R. Sunstein, Is Deontology a Heuristic? On Psychology, Neuroscience, Ethics, and Law (Sept. 1, 2013).

Published Version

http://www.radcliffe.harvard.edu/exploratory-seminars/public-policy-and-brain-supplemental-resources

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:13548959

Terms of Use

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Is Deontology A Heuristic? On Psychology, Neuroscience, Ethics, and Law

Cass R. Sunstein[*]

Abstract

A growing body of psychological and neuroscientific research links dual-process theories of cognition with moral reasoning (and implicitly to legal reasoning as well). The relevant research appears to show that at least some deontological judgments are connected with rapid, automatic, emotional processing, and that consequentialist judgments (including utilitarianism) are connected with slower, more deliberative thinking. These findings are consistent with the claim that deontological thinking is best understood as a moral heuristic – one that generally works well, but that also misfires. If this claim is right, it may have large implications for many debates in politics, morality, and law, including those involving the role of retribution, the free speech principle, religious liberty, the idea of fairness, and the legitimacy of cost-benefit analysis. Nonetheless, psychological and neuroscientific research cannot rule out the possibility that consequentialism is wrong and that deontology is right. It tells us about the psychology of moral and legal judgment, but it does no more. On the largest questions, it leaves moral and legal debates essentially as they were before.

I. Deontology As Heuristic

A. The Problem

In political, moral, and legal theory, many of the largest debates pit consequentialists against deontologists. Consequentialists believe that the rightness of actions turns on their consequences, which are to be measured, aggregated, and ranked.[1] (Utilitarianism is a species of consequentialism.) By contrast, deontologists believe that some actions are wrong even if they have good consequences. On this view, it is

---

[1] *See* MATTHEW ADLER, WELL-BEING AND FAIR DISTRIBUTION: BEYOND COST-BENEFIT ANALYSIS (2011).

generally wrong to torture people, or to kill them, even if the consequences of doing so would be good. You should not throw someone in the way of a speeding train even if that action would save lives on balance; you should not torture someone even if doing so would produce information that would save lives; slavery is a moral wrong regardless of the outcome of any utilitarian calculus; the protection of free speech does not depend on any such calculus; the strongest arguments for and against capital punishment turn on what is right, independent of the consequences of capital punishment.[2]

The disagreements between deontologists and consequentialists bear directly on many issues in law and policy. For example, consequentialists favor theories of punishment that are based on deterrence, and they firmly reject retributivism, which many deontologists endorse.[3] For both criminal punishment and punitive damage awards, consequentialists and deontologists have systematic disagreements. Consequentialists believe that constitutional rights (including the right to free speech[4]) must be defended and interpreted by reference to the consequences; deontologists disagree. Consequentialists are favorably disposed to cost-benefit analysis in regulatory policy,[5] but that form of analysis has been vigorously challenged on broadly deontological grounds.[6] Consequentialists and deontologists also disagree about the principles underlying the law of contract[7] and tort.[8]

In defending their views, deontologists often point to cases in which our intuitions seem very firm, and hence to operate as "fixed points" against which we must evaluate consequentialism.[9] They attempt to show that consequentialism runs up against intransigent intuitions and is wrong for that reason.[10] And indeed it is true that consequentialism seems to conflict with such of our deepest intuitions, certainly in new or unfamiliar situations. For example, human beings appear to be intuitive retributivists, and with respect to punishment, efforts to encourage people to think in consequentialist terms do not fare at all well.[11]

---

[2] A helpful discussion is Mark Timmons, *Toward A Sentimentalist Deontology*, *in* MORAL PSYCHOLOGY, VOL. 3: THE NEUROSCIENCE OF MORALITY: EMOTION, BRAIN DISORDERS, AND DEVELOPMENT (Walter-Sinnott-Armstrong ed., 2007).

[3] *See* Dan Markel, *Retributive Damages*, 94 CORN. L. REV. 239, 243, 254-57 (2009).

[4] *See* Richard A. Posner, *Free Speech in An Economic Perspective*, 20 SUFFOLK L. REV. 1, 8 (2000).

[5] *See* W. KIP VISCUSI, FATAL TRADEOFFS (1991).

[6] *See* FRANK ACKERMAN AND LISA HEINZERLING, PRICELESS (2006); DOUGLAS KYSAR, REGULATING FROM NOWHERE (2011).

[7] *See* CHARLES FRIED, CONTRACT AS PROMISE (1982).

[8] *See* Kenneth W. Simons *Tort Negligence, Cost-Benefit Analysis, and Tradeoffs: A Closer Look at the Controversy*, 41 LOYOLA OF LOS ANGELES L. REV. 1171, 1172-73 (2007-08).

[9] *See, e.g.*, FRANCES KAMM, INTRICATE ETHICS (2006), for a recent discussion; a classic treatment is Bernard Williams, *A Critique of Utilitarianism*, in J.C. Smart and Bernard Williams, Utilitarianism: For and Against (1973).

[10] A version of this strategy is used in MICHAEL SANDEL, WHAT MONEY CAN'T BUY (2011).

[11] *See* Cass R. Sunstein et al., *Do People Want Optimal Deterrence?*, 29 J. LEGAL STUD. 237, 250 (2000).

In the face of the extensive body of philosophical work exploring the conflict between deontology and consequentialism, it seems reckless to venture a simple resolution, but let us consider one: Deontology is a *moral heuristic* for what really matters, and consequences are what really matter. On this view, deontological intuitions are generally sound, in the sense that they usually lead to what would emerge from a proper consequentialist assessment. Protection of free speech and religious liberty, for example, is generally justified on consequentialist grounds. At the same time, however, deontological intuitions can sometimes produce severe and systematic errors, in the form of suboptimal or bad consequences.[12] The idea that deontology should be seen as a heuristic is consistent with a growing body of behavioral and neuroscientific research, which generally finds that deontological judgments are rooted in automatic, emotional processing.[13]

My basic claims here are twofold. The first is that the emerging research might serve to unsettle and loosen some deeply held moral intuitions,[14] and give us new reason to scrutinize our immediate and seemingly firm reactions to moral problems. The second is that nothing in that research is sufficient to establish that deontological judgments are wrong or false. Deontology may in fact be a moral heuristic,[15] in the sense that it may be a mental shortcut for the right moral analysis, which is consequentialist.[16] But the neuroscientific and psychological research is neither necessary nor sufficient to establish this claim. What is required is not a psychology of moral argument, but a moral argument.[17]

II. Intuitions About Fact, Intuitions About Value

---

[12] This point holds, for example, in the context of punitive damage awards. See Sunstein et al., supra note.

[13] See below. Note, however, that the relevant findings cover only a very small portion of the domain in which deontological judgments are and might be made, and that some evidence does not support the view that such judgments are distinctly associated with automatic processing. Andrea Manfrinati et al., *Moral Dilemmas and Moral Principles: When Emotion and Cognition Unite*, Cognition & Emotion (2013). Note also that learning and culture certainly matter, and we do not have much cross-cultural evidence, which would be highly informative about the relationships among deontology, automatic processing, and culture.

[14] For a valuable discussion, *see generally* Mark Kelman, *Intuitions*, 65 STAN. L. REV. 1291 (2013).

[15] *See* Cass R. Sunstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCIENCES 531, 531 (2005). In that essay, I explore the possibility that certain mid-level principles, in law and morality, should be counted as moral heuristics, while bracketing the larger question of whether deontology itself can be so understood -- the question engaged here.

[16] For an elaborate discussion of the relationship between the two, see Derek Parfitt, On What Matters, vol. 1 (2011).

[17] A broadly similar conclusion is defended on somewhat different grounds in Selim Berker, *The Normative Insignificance of Neuroscience*, 37 Phil & Public Affairs 293 (2009). To the same general effect, see F.M. Kamm, *Neuroscience and Moral Reasoning: A Note on Recent Research*, 37 Phil& Public Affairs 330 (2009). For a clear, concise overview, see Arleen Salles, *On the Normative Implications of Social Neuroscience*, 13 Recerca Revista De Pensament I Analisi 29 (2013).

A. An Analogy

If deontology is a heuristic, then deontological intuitions may be functionally equivalent to those elicited by Daniel Kahneman and Amos Tversky in their famous experiments on judgment and decision-making.[18] Kahneman and Tversky showed that in answering hard factual questions, people use simple rules of thumb.[19] Consider the representativeness heuristic, in accordance with which judgments of probability are influenced by assessments of resemblance (the extent to which A "looks like" B).[20] The representativeness heuristic is famously exemplified by people's answers to questions about the likely career of a hypothetical woman named Linda, described as follows: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations." People were asked to rank, in order of probability, eight possible futures for Linda. Six of these were fillers (such as psychiatric social worker, elementary school teacher); the two crucial ones were "bank teller" and "bank teller and active in the feminist movement."

Most people say that Linda was less likely to be a bank teller than a bank teller and active in the feminist movement. This is an obvious mistake, a conjunction error, in which characteristics A and B are thought to be more likely than characteristic A alone. The error stems from the representativeness heuristic: Linda's description seems to match "bank teller and active in the feminist movement" far better than "bank teller." In an illuminating and even profound reflection on the example, Stephen Jay Gould observes that "I know [the right answer], yet a little homunculus in my head continues to jump up and down, shouting at me – 'but she can't just be a bank teller; read the description.'"[21]

---

[18] The principal paper is Amos Tversky and Daniel Kahneman, Judgment Under Uncertainty: Heuristics and Biases, 185 Science 1124 (1974); it is reprinted as an appendix to DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2010). For collections, see Judgment Under Uncertainty: Heuristics and Biases (Daniel Kahneman et al. eds. 1982); Heuristics and Biases: The Psychology of Intuitive Judgment (Thomas Gilovich et al. eds. 2002). I draw in the next several paragraphs on Sunstein, *Moral Heuristics*, *supra* note.

[19] *See* KAHNEMAN, *supra* note; for an explanation of how heuristics generally work well, see GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (2005). There is considerable tension between the work of Kahneman and Tversky on the one hand and that of Gigerenzer and his colleagues on the other. I do not discuss that tension here, though it has implications for the current topic. For example, Gigerenzer has explored the idea that ordinary morality consists of moral heuristics, which (in his view) work quite well. See Gerd Gigerenzer, *Moral Intuition = Fast and Frugal Heuristics?*, in Moral Psychology, vol. 2: The Cognitive Sciences of Morality: Intuition and Diversity 1 (Walter Sinnott-Armstrong ed. 2008). For a different view, see Cass R. Sunstein, *Fast, Frugal, and (Sometimes) Wrong*, in id. at 27. An illuminating discussion is MARK KELMAN, THE HEURISTICS DEBATE (2011).

[20] The best discussion, emphasizing that those who use heuristics substitute an easy question for a hard one ("attribute substitution"), is Daniel Kahneman and Shane Frederick, *Representativeness Revisited*, in Heuristics and Biases, supra note.

[21] STEPHEN JAY GOULD, BULLY FOR BRONTOSAURUS: REFLECTIONS IN NATURAL HISTORY 469 (1991).

(We should keep Gould's homunculus in mind; because of its relevance for moral issues, I will return to it several times.)

Or consider these questions: How many words, in four pages of a novel, will have "ing" as the last three letters? How many words, in the same four pages, will have "n" as the second-to-last letter? Most people will give a higher number in response to the first question than in response to the second[22] – even though a moment's reflection shows that this is a mistake. People err because they use an identifiable heuristic – the availability heuristic – to answer difficult questions about probability.[23] When people use this heuristic, they answer a question about probability by asking whether examples come readily to mind. How likely is a flood, an airplane crash, a traffic jam, a terrorist attack, or a disaster at a nuclear power plant? To answer such questions, people try to think of illustrations. For people without statistical knowledge, it is far from irrational to use the availability heuristic. When people enlist heuristics, they substitute an easy question for a hard one, and use of the representativeness and availability heuristics can be understood and defended in that light.[24] The problem is that both heuristics can lead to serious errors of fact. With the availability heuristic, those errors take the form of excessive fear of small risks and neglect of large ones.

"Availability bias" is the source of many mistakes about facts. There is an independent standard for the truth, and the judgments of those who use the availability heuristic are sometimes wrong. With respect to morality, is it possible to speak of mistakes? If so, can it be said that "deontology bias" is the source of many mistakes about morality? If the answer to both questions is "yes," then a serious question might be raised about the idea of seeking "reflective equilibrium," elaborated by John Rawls, which suggests that we sort out moral questions by seeking some kind of equilibrium among the various views that we hold.[25] Some of our moral intuitions seem very firm (Gould's homunculus swears by them), and their firmness may be a good thing (because it has good consequences), but perhaps they misfire, and perhaps we should be willing to reconsider them, at least if the result would be to produce more accuracy (in the form of better consequences) than can result from a mere heuristic. Thus Henry Sidgwick urges[26]:

> It may be shown, I think, that the Utilitarian estimate of consequences not only supports broadly the current moral rules, but also sustains their generally received limitations and qualifications: that, again, it explains

---

[22] Amos Tversky and Daniel Kahneman, *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, 90 PSYCHOLOGICAL REVIEW 293, 295 (1984).

[23] Amos Tversky and Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 Cognitive Psychology 207 (1973). For applications in the domains of politics and law, see Timur Kuran and Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 Stan. L. Rev. 683 (1999).

[24] See Kahneman and Frederick, supra note.

[25] *See* JOHN RAWLS, A THEORY OF JUSTICE (1971). That question is pressed explicitly in Peter Singer, Ethics and Intuitions, 9 Journal of Ethics 331 (2005).

[26] HENRY SIDGWICK, THE METHODS OF ETHICS 425-26 (1981 ed.).

anomalies in the Morality of Common Sense, which from any other point of view must seem unsatisfactory to the reflective intellect; and moreover, where the current formula is not sufficiently precise for the guidance of conduct, while at the same time difficulties and perplexities arise in the attempt to give it additional precision, the Utilitarian method solves these difficulties and perplexities in general accordance with the vague instincts of Common Sense, and is naturally appealed to for such solution in ordinary moral discussions.

Sidgwick did not have the benefit of recent empirical findings, but he might be willing to agree "that deontological philosophy is largely a rationalization of emotional moral intuitions."[27]

B. System 1 and System 2

Within social science, it has become standard to suggest that the human mind contains not one but two "cognitive systems."[28] In the social science literature, the two systems are unimaginatively described as System 1 and System 2.[29] System 1 is the automatic system, while System 2 is more deliberative and reflective.

System 1 works fast. It relies on intuitions, and it can be emotional. Much of the time, it is on automatic pilot. When it hears a loud noise, it is inclined to run. When it is offended, it wants to hit back (hence it is retributive). It certainly eats a delicious brownie. It can procrastinate; it can be impulsive. It wants what it wants when it wants it. It can be excessively fearful and too complacent. It is a doer, not a planner. System 1 is a bit like Homer Simpson, James Dean (from *Rebel Without A Cause*), and Pippi Longstocking.

System 2 is more like a computer or Mr. Spock from the old *Star Trek* show. It is deliberative. It calculates. It hears a loud noise, and it assesses whether the noise is a cause for concern. It thinks about probability, carefully though sometimes slowly. If it sees reasons for offense, it makes a careful assessment of what, all things considered, ought to be done. It insists on the importance of self-control. It is a planner as well as a doer; it does what it has planned.

At this point, it might be asked: What, exactly, are these systems? Do they operate as homunculi in the brain? The best answer is that the idea of two systems is a heuristic device, a simplification that is designed to refer to automatic, effortless processing and more complex, effortful processing.[30] When people are asked to add one plus one, or to

---

[27] Joshua D. Greene, *Reply to Mikhail and Timmons*, *in* MORAL PSYCHOLOGY, VOL. 3: THE NEUROSCIENCE OF MORALITY: EMOTION, BRAIN DISORDERS, AND DEVELOPMENT (Walter-Sinnott-Armstrong ed., 2007).

[28] *See* KAHNEMAN, *supra* note, for an instructive discussion of the distinction, which psychologists have used for over a decade.

[29] *See id.* at 13.

[30] *See id.*

walk from their bedroom to their bathroom in the dark, or to read the emotion on the face of their best friend, the mental operation is easy and rapid. When people are asked to multiply 179 times 283, or to navigate a new neighborhood by car, or to decide which retirement or health insurance plan best fits their needs, the mental operation is difficult and slow.

Identifiable regions of the brain are active in different tasks, and hence it may well be right to suggest that the idea of "systems" has physical referents. An influential discussion states that "[a]utomatic and controlled processes can be roughly distinguished by where they occur in the brain."[31] The prefrontal cortex, the most advanced part of the brain (in terms of evolution) and the part that most separates human beings from other species, is associated with deliberation and hence with System 2. The amygdala has been associated with a number of automatic processes, including fear,[32] and it can thus be associated with System 1. On the other hand, different parts of the brain interact, and it would be hazardous to locate System 1 and System 2 in different regions. There is no need to make technical or controversial claims about neuroscience in order to distinguish between effortless and effortful processing. The idea of System 1 and System 2 is designed to capture that distinction in a way that works for purposes of exposition (and that can be grasped fairly immediately by System 1).

Many cognitive errors, emphasized by behavioral scientists, reflect the primacy of System 1. Confronted with a difficult question, people answer an easier one. Perhaps the same is true in the moral domain. System 1 might produce quick, intuitive answers to moral questions. On the view that I am considering here, deontological approaches should be taken to reflect the power of System 1, whereas consequentialist approaches reflect the work of System 2. Note that this is a point about moral psychology; it is not a suggestion that System 2 is unable to produce elaborate and plausible defenses of deontology. But if deontology is a heuristic, we should consider the possibility that in providing those defenses, System 2 is acting as System 1's lawyer, or public relations manager, and that System 1 is in charge, at least initially and perhaps throughout.[33]

---

[31] Colin Camerer et al., *Neuroeconomics: How Neuroscience Can Inform Economics*, 43 J. ECON. LITERATURE 1, 17 (2005).

[32] *See* THE HUMAN AMYGDALA (Paul J. Whalen & Elizabeth A. Phelps eds., 2009). At the same, time, the subcortical systems (including the amygdala) are in fact capable of highly sophisticated calculations and processes. Human beings can learn a great deal using the amygdala and striatum. See Li Schiller et al., *Differential roles of human striatum and amygdala in associative learning*, 11 Nature Neuroscience 1250 (2011). The frontal lobes, on the other hand, may turn out to be the biasing source. See, e.g., Tali Sharot et al., *How Unrealistic Optimism Is Maintained In The Face of Reality*, 14 Nature Neuroscience 1475 (2011).

[33] On the general point, see Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 Psychological Review 814 (2001). Some moral judgments do seem to reflect conclusions that antedate reasons and that have everything to do with intuitive, visceral reactions, not with reasons at all; and some of those judgments seem broadly deontological. See Jonathan Haidt et al., *Moral Dumbfounding: When Intuition Finds No Reason* (2000), available at http://commonsenseatheism.com/wp-content/uploads/2011/08/Haidt-Moral-Dumfounding-When-Intuition-Finds-No-Reason.pdf.

III. Neuroscience, Behavior, and Deontology

A. Trolleys and Footbridges

A great deal of neuroscientific and psychological work is consistent with the view that deontological judgments stem from a moral heuristic, one that works automatically and rapidly.[34] It bears emphasizing that if this view is correct, it is also possible, indeed likely, that such judgments generally work well, in the sense that they produce the right results (according to the appropriate standard) in most cases. The judgments that emerge from automatic processing, including emotional varieties, usually turn out the way they do for a reason.[35] If deontological judgments result from a moral heuristic, we may end up concluding that they generally work well, but that they misfire in systematic ways.[36]

Consider in this regard the longstanding philosophical debate over two well-known moral dilemmas, which seem to test deontology and consequentialism.[37] The first, called the trolley problem, asks people to imagine that a runaway trolley is headed for five people, who will be killed if the trolley continues on its current course. The question is whether you would throw a switch that would move the trolley onto another set of tracks, killing one person rather than five. Most people would throw the switch. The second, called the footbridge problem, is the same as that just given, but with one difference: the only way to save the five is to throw a stranger, now on a footbridge that spans the tracks, into the path of the trolley, killing that stranger but preventing the trolley from reaching the others. Most people will not kill the stranger.

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, or rescued, as a matter of principle.[38] The basic idea seems to be that those firm intuitions, separating the two cases, tell us something important about what morality requires, and an important philosophical task is to explain why they are essentially right.

Without engaging these arguments, consider a simpler answer. As a matter of principle, there is no difference between the two cases. People's different reactions are based on a deontological heuristic ("do not throw innocent people to their death") that

---

[34] For contrary evidence, however, see Andrea Manfrinati et al., *Moral Dilemmas and Moral Principles: When Emotion and Cognition Unite*, Cognition & Emotion (2013).

[35] Note also that culture is especially important in producing those judgments, even if they are automatic. See Haidt, supra note. It follows that in different cultures, automatic processing could well lead to quite different judgments.

[36] This is consistent with a longstanding view in the utilitarian tradition and is developed at length in STEVEN SHAVELL AND LOUIS KAPLOW, FAIRNESS VERSUS WELFARE (2004).

[37] Judith Jarvis Thomson, *The Trolley Problem. in* RIGHTS, RESTITUTION AND RISK: ESSAYS IN MORAL THEORY (J.J. Thomson & W. Parent, eds. 1986).

[38] *See, e.g.*, *id.;* F.M. Kamm, Intricate Ethics (2006). For a criticism of the use of hypothetical problems of this sort, see James O'Connor, *The Trolley Method of Moral Philosophy*, 13 Essays in Philosophy (2012), available at http://commons.pacificu.edu/cgi/viewcontent.cgi?article=1413&context=eip

condemns the throwing of the stranger but not the throwing of the switch. To say the least, it is desirable for people to act on the basis of a moral heuristic that makes it extremely abhorrent to use physical force to kill innocent people. But the underlying heuristic misfires in drawing a distinction between the two ingeniously devised cases. Hence people (including philosophers) struggle heroically to rescue their intuitions and to establish that the two cases are genuinely different in principle. But they are not.[39] If so, a deontological intuition is serving as a heuristic in the footbridge problem, and it is leading people in the wrong direction. Can this proposition be tested? Does it suggest something more general about deontology?

A. Neuroscience

1. Brains, trolleys, and footbridges. How does the human brain respond to the trolley and footbridge problems[40]? The authors of an influential study 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," and that brain areas associated with emotion are far more active in contemplating the footbridge problem than in contemplating the trolley problem.[41]

More particularly, the footbridge problem preferentially activates the regions of the brain that are associated with emotion, including the amygdala, the mPFC, and the PCC. By contrast, the trolley problem produces increased activity in parts of the brain associated with cognitive control and working memory. A possible implication of the authors' finding is that human brains distinguish between different ways of bringing about deaths; some ways trigger automatic, emotional reactions, while others do not. Other fMRI studies reach the same general conclusion.[42]

2. Actions and omissions. People tend to believe that harmful actions are worse than harmful omissions; intuition does suggest a sharp difference between the two. Many people believe that the distinction is justified in principle.[43] They may be right, and the arguments offered in defense of the distinction might be convincing; but in terms of people's actual judgments, there is reason to believe that automatic (as opposed to deliberative or controlled) mechanisms help to account for people's intuitions.[44] Controlled cognition is indicated by the frontoparietal control network and especially the

---

[39] See Singer, supra note, at 347-50.

[40] Greene, J., Somerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. *An fMRI Investigation of Emotional Engagement in Moral Judgment*, 293 SCIENCE 2105, 2106 (2001). Various questions are raised about the methodology here in Berker, supra note, at 305-313; I bracket those questions here.

[41] Greene et al., supra. On the limits of fMRI studies, and in particular the difficulty of inferring causation, see note infra.

[42] *See* Joshua Greene, *The Cognitive Neuroscience of Moral Judgment*, *in* THE COGNITIVE NEUROSCIENCES 4TH ED. (Michael S. Gazzaniga ed., 2009).

[43] See John Hall, *Acts and Omissions*, 39 The Philosophical Quarterly 199 (1989).

[44] *See* Fiery Cushman et al., *Judgment Before Principle: Engagement of the Frontoparietal Control Network*, *in* CONDEMNING HARMS OF OMISSION, SOCIAL COGNITIVE AND AFFECTIVE NEUROSCIENCE (2011).

dorsolateral prefrontal cortex. The dorsal medial prefrontal cortex (DMPFC) plays an essential role in controlled reasoning.

In the relevant experiments, participants were presented with a series of moral scenarios, involving both actions and omissions. They judged active harms to be far more objectionable than inaction. As compared with harmful actions, harmful omissions produced significantly more engagement in the frontoparietal control network. Those participants who showed the *highest* engagement in that network, while answering questions involving omissions, also tended to show the *smallest* differences in their judgments of actions and omissions. This finding suggests that a high level of controlled processing was necessary to override the intuitive sense that the two are different (with omissions seeming less troublesome), and hence that there is "a role for controlled cognition in the elimination of the intuition effect."[45] The upshot is that lesser concern with omissions arises automatically, without the use of controlled cognition. Such cognition is apparently used to overcome automatic judgment processes in order to condemn harmful omissions. Hence "controlled cognition is associated not with conforming to the omission effect but with overriding it,"[46] and "the more a person judges harmful omissions on parity with harmful actions, the more they engage cognitive control during the judgment of omissions."[47]

3. Social Emotions and Utilitarianism. The ventromedial prefrontal cortex (VMPC) is a region of the brain that is necessary for social emotions, such as compassion, shame, and guilt.[48] Patients with VMPC lesions show reductions in these emotions and reduced emotional receptivity in general. Researchers predicted that such patients would show an unusually high rate of utilitarian judgments in moral scenarios that typically trigger strong emotions – such as pushing someone off a bridge to prevent a runaway boxcar from hitting five people. The prediction turned out to be correct.[49] Those with damage to the VMPC engaged in utilitarian reasoning in responding to such problems.

This finding is consistent with the view that deontological reasoning is a product of negative emotional responses that predict moral disapproval.[50] By contrast, consequentialist reasoning, reflecting a kind of cost-benefit analysis, is subserved by the dorsolateral prefrontal cortex. Damage to the VMPC predictably dampens the effects of emotions and leads people to engage in an analysis of likely effects of different courses of action. Similarly, people with frontotemporal dementia are believed to suffer from "emotional blunting" – and they are especially likely to favor action in the footbridge

---

[45] *Id.*

[46] *Id.*

[47] *Id.*

[48] *See* Michael Koenigs et al., *Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments*, 446 NATURE 908, 908 (2007).

[49] *See id.* at 909-10.

[50] Joshua Greene, *Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains*, 11 TRENDS IN COGNITIVE SCI. 322, 322 (2007).

problem.[51] On an admittedly controversial interpretation of these findings, "patients with emotional deficits may, in some contexts, be the most pro-social of all."[52]

C. Behavioral Evidence and Deontology

A great deal of behavioral evidence also suggests that deontological thinking is associated with System 1 and in particular with emotions.

1. Words or pictures? People were tested to see if they had a visual or verbal "cognitive style," that is, to see whether they performed better with tests of visual accuracy than with tests of verbal accuracy.[53] The authors hypothesized that because visual representations are more emotionally salient, those who do best with verbal processing would be more likely to support utilitarian judgments, and those who do best with visual processing would be more likely to support deontological judgments. The hypothesis was confirmed. Those with more visual cognitive styles were more likely to favor deontological approaches.[54]

People's self-reports showed that their internal imagery – that is, what they visualized – predicted their judgments, in the sense that those who "saw" pictures of concrete harm were significantly more likely to favor deontological approaches. In the authors' words, "visual imagery plays an important role in triggering the automatic emotional responses that support deontological judgments."[55]

2. The effects of cognitive load. What are the effects of "cognitive load"? If people are asked to engage in tasks that are cognitively difficult, such that they have less "space" for complex processing, what happens to their moral judgments? The answer is clear: An increase in cognitive load interferes with consequentialist (utilitarian) moral judgment but has no such effect on deontological approaches.[56] This finding strongly supports the view that consequentialist judgments are cognitively demanding and that deontological judgments are relatively easy and automatic.[57]

---

[51] Mendez et al., *An Investigation of Moral Judgement in Frontotemporal Dementia*, 18 COGNITIVE & BEHAVIORAL NEUROLOGY 193 (2005).

[52] *Id.* Note that on their own, fMRI studies suggest only correlations and cannot distinguish cause and effect. If region X is active when we make decision Y, it is not necessarily the case that X is causing decision Y. Y may be causing X, or both may be caused by something else all together. For example, the act of making a deontological judgment may cause an emotional reaction that may be processed by the amygdala and/or vmpfc. By contrast, lesions studies may suggest cause and effect. (I am grateful to Tali Sharot for clarifying this point.)

[53] *See* Elinor Amit and Joshua Greene, *You See, The Ends Don't Justify the Means: Visual Imagery and Moral Judgment*, 23 PSYCH. SCI. 861, 862 (2012).

[54] *Id.*

[55] *Id.* at 866.

[56] Joshua D. Greene, *Cognitive Load Selectively Interferes With Utilitarian Moral Judgment*, 107 COGNITION 1144, 1151 (2008).

[57] Contrary evidence, suggesting that deontological thinking can actually take longer than consequentialist thinking, and that "cognitive and emotional processes participate in both deontological and consequentialist moral judgments," can be found in Andrea Manfrinati et al.,

3. *Priming System 2.* The "Cognitive Reflection Test" (CRT) asks a series of questions that elicit answers that fit with people's intuitions but that turn out to be wrong. Here is one such question: *A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?* In response, most people do not give the correct answer, which is 5 cents. They are more likely to offer the intuitively plausible answer, which is 10 cents. Those who take the CRT tend to learn that they often give an immediate answer that turns out, on reflection, to be incorrect. If people take the CRT before engaging in some other task, they will be "primed" to question their own intuitive judgments. What is the effect of taking the CRT on moral judgments?

The answer is clear: Those who take the CRT are more likely to reject deontological thinking in favor of utilitarianism.[58] Consider the following dilemma:

John is the captain of a military submarine traveling underneath a large iceberg. An onboard explosion has caused the vessel to lose most of its oxygen supply and has injured a crewman who is quickly losing blood. The injured crewman is going to die from his wounds no matter what happens.

The remaining oxygen is not sufficient for the entire crew to make it to the surface. The only way to save the other crew members is for John to shoot dead the injured crewman so that there will be just enough oxygen for the rest of the crew to survive.

Is it morally acceptable for John to shoot the injured crewman? Those who took the CRT first, before answering that question, were far more likely to find that action morally acceptable.[59] Across a series of questions, those who took the CRT became significantly more likely to support consequentialist approaches to social dilemmas.

IV. Morality and the Need for Independent Standards

A. Is and Ought

The evidence just outlined is consistent with the proposition that deontological intuitions are mere heuristics, produced by the automatic operations of System 1.[60] The basic picture would be closely akin to the corresponding one for questions of fact. People use mental shortcuts, or rules of thumb, that generally work well, but that can also lead to systematic errors.

Moral Dilemmas and Moral Principles: When Emotion and Cognition Unite, Cognition & Emotion (2013).

[58] *See* Joseph Paxton et al., *Reflection and Reasoning in Moral Judgment*, 36 COGNITIVE SCI. 163, 171-72 (2011).

[59] *Id.* at 166.

[60] For contrary evidence, however, see Andrea Manfrinati et al., Moral Dilemmas and Moral Principles: When Emotion and Cognition Unite, Cognition & Emotion (2013).

To be sure, the neuroscientific and psychological evidence is preliminary and suggestive but no more. Importantly, we do not have much cross-cultural evidence. Do people in diverse nations and cultures show the same kinds of reactions to moral dilemmas? Is automatic processing associated with deontological approaches only in certain nations and cultures? Do people in some nations show automatic moral disapproval (perhaps motivated by disgust) of practices and judgments that seem tolerable or even appropriate elsewhere[61]? Where and how does deontology matter? There may not be simple answers to such questions; perhaps some deontological reactions are hard-wired and others are not.

Moreover, deontological intuitions and judgments span an exceedingly wide range. They are hardly exhausted by the trolley and footbridge problems and by related hypothetical questions, and if deontological intuitions are confused or unhelpful in resolving such problems, deontology would not stand defeated by virtue of that fact. Consider, for example, retributive theories of punishment[62]; autonomy-based theories of freedom of speech and religion; bans on slavery and torture, grounded in principles of respect for persons; and theories of tort law and contract law that are rooted in conceptions of fairness.[63] We do not have neuroscientific or psychological evidence with respect to the nature and role of deontological thinking in the wide assortment of moral, political, and legal problems for which deontological approaches have been proposed or defended. Perhaps System 2, and not System 1, is responsible for deontological thinking with respect to some of those problems. It is certainly imaginable, however, that neuroscientific or psychological evidence will eventually find that automatic processing supports deontological thinking across a wide range of problems.

1. *The central objection*. Even if this is so, the proposition that deontology is a heuristic (in the sense in which I am using that term) runs into a serious and immediate objection. For factual matters, we have an independent standard by which to assess the question of truth. Suppose that people think that more people die as a result of homicide than suicide. The facts show that people's judgments, influenced by the availability heuristic, are incorrect. But if people believe that torture is immoral even if it has good consequences, we do not have a self-evidently correct independent standard to demonstrate that they are wrong.

To be sure, a great deal of philosophical work attempts to defend some version of consequentialism.[64] But deontologists reject the relevant arguments. They do so for what they take to be good reasons, and they elaborate those reasons in great detail. With respect to facts, social scientists can show that certain rules-of-thumb produce errors; the same cannot be said for deontology. For this reason, deontological judgments may not a

---

[61] For reason to suspect an affirmative answer, see Jonathan Haidt, The Righteous Mind (2012).

[62] For discussion, see Michael S. Pardo and Dennis Patterson, Neuroscientific Challenges to Retributivism, in The Future of Punishment (Thomas Nadelhoffer ed., forthcoming 2013).

[63] See STEVEN SHAVELL AND LOUIS KAPLOW, FAIRNESS VERSUS WELFARE (2004).

[64] For a valuable recent discussion, see Matthew D. Adler, Well-Being and Fair Distribution (2011).

mental shortcut at all. Even if automatic processing gives them a head start, they may ultimately be the product of a long and successful journey.

Or suppose that certain moral intuitions arise in part because of emotions and that some or many deontological intuitions fall in that category. Even if so, we would not have sufficient reason to believe that those intuitions are wrong.[65] Some intuitions about states of the world arise from the emotion of fear, and they are not wrong for that reason. To be sure, people may be fearful when they are actually safe, but without knowing about the situations that cause fear, we have no reason to think that the emotion is leading people to make mistakes. The fact – if it is a fact[66] – that some or many deontological intuitions are driven by emotions does not mean that those intuitions misfire.

If these points are right, then we might be able to agree that for many people, deontological thinking often emerges from automatic processing and that consequential thinking is often more calculative and deliberative. This might well be the right way to think about moral psychology, at least in many domains, and the resulting understanding of moral psychology certainly has explanatory power for many problems in law and politics; it helps us to understand why legal and political debates take the form that they do.[67] But if so, we would not be able to conclude that deontological thinking is wrong.[68] Consider in this regard the fact that in response to some problems and situations, people's immediate, intuitive responses are right, and a great deal of reflection and deliberation can produce mistakes.[69] There is no reason to think that System 2 is always more accurate than System 1. Even if deontological judgments are automatic and emotional, they may turn out to be correct.[70]

2. *Two new species*. Here is a way to sharpen the point. Imagine that we discovered two new species of human beings: Kantians and Benthamites. Suppose that the Kantians are far more emotional than home sapiens and that Benthamites are far less so. Imagine that neuroscientists learn that Kantians and Benthamites have distinctive brain structures. Kantians have a highly developed emotional system and a relatively undeveloped cognitive system. By contrast, Benthamites have a highly developed cognitive system – significantly more developed than that in homo sapiens. And true to their names, Kantians strongly favor deontological approaches to moral questions, while Benthamites are thoroughgoing consequentialists.

Impressed by this evidence, some people insist that we have new reason to think that consequentialism is correct. And indeed, anthropologists discover that Benthamites have written many impressive and elaborate arguments in favor of consequentialism. By

---

[65] See Berker, supra note, at 316-317.

[66] For a more complex set of findings, see Andrea Manfrinati et al., Moral Dilemmas and Moral Principles: When Emotion and Cognition Unite, Cognition & Emotion (2013).

[67] It might help explain, for example, why some forms of economic and utilitarian thinking run into serious objections and sometimes even incredulity in the political and legal domains.

[68] To the same effect, *see* Timmons, *supra* note.

[69] See Sian Beilock, Choke (2009).

[70] For many cautionary notes, see Haidt, supra note.

contrast, Kantians have written nothing. (They do not write much.) With such discoveries, would we have new reason to think that consequentialism is right? The answer would seem to turn on the strength of the arguments offered on behalf of consequentialism, not on anything about the brains of the two species.[71]

Now suppose that an iconoclastic Benthamite has written a powerful essay, contending that consequentialism is wrong and that some version of Kantianism is right. Wouldn't that argument have to be investigated on its merits[72]? If the answer is affirmative, then we should be able to see that even if certain moral convictions originate in automatic processing, they may nonetheless be correct. Everything depends on the justifications that have been provided in their defense. A deontological conviction may come from System 1, but the Kantians might be right, and the Benthamites should listen to what they have to say.[73]

B. Moral Reasoning and Moral Rationalization

Suppose that we agree that recent research shows that as a matter of fact, "deontological judgments tend to be driven by emotional responses"; a more provocative conclusion, consistent with (but not mandated by) the evidence, is that "deontological philosophy, rather than being grounded in moral *reasoning*, is to a large extent an exercise in moral *rationalization*."[74] Without denying the possibility that the intuitive system is right, Greene contends "that science, and neuroscience in particular, can have profound ethical implications by providing us with information that will prompt us to re-evaluate our moral values and our conceptions of morality."[75]

The claim seems plausible. But how, exactly, might scientific information prompt us to re-evaluate our moral values? The best answer is that it might lead people to slow

---

[71] Compare Kamm, supra note, emphasizing the existence of gaps in "an argument from fMRI data to a normative conclusion. The argument sounds like a pure procedural argument for rightness with the brain part substituting for a procedure; if a particular procedure (or brain part) leads to an answer, it is the right answer. Those who rely on a procedural account of rightness, however, must give an independent argument for what makes this the right procedure to use."

[72] *Cf.* Greene, *supra* note: "Couldn't deontological thinking be the right kind of thinking, regardless of our psychological motivates for embracing it? It could." Greene is of course correct here; note his further suggestion that this is "the deontological dream" and that "keeping that particular dream alive requires one to posit a strange set of coincidences . . . ."

[73] Berker, supra note, at 329, explores the idea that we might discount certain moral intuitions if "[w]e notice that a portion of the brain which lights up whenever we make a certain sort of obvious, egregious error in mathematical or logical reasoning also lights up whenever we have a certain intuition."

[74] *See* Joshua Green, *The Secret Joke of Kant's Soul*, *in* MORAL PSYCHOLOGY, VOL. 3: THE NEUROSCIENCE OF MORALITY: EMOTION, BRAIN DISORDERS, AND DEVELOPMENT (Walter-Sinnott-Armstrong ed., 2007).

[75] Joshua Greene, *From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?*, 4 NATURE REVIEWS NEUROSCIENCE 847, 847 (2003). For a critical discussion of the rationalization claim, see Richard Dean, *Does Neuroscience Undermine Deontological Theory*, 3 Neuroethics 43 (2010).

down and to give real scrutiny to their immediate reactions. If you know that your moral judgment is a rapid intuition, based on emotional processing, you might be more willing to consider the possibility that it is wrong. You might be willing to consider the possibility that you have been influenced by irrelevant factors.

Suppose that you believe that it is unacceptable to push someone into a speeding train even if you know that the result of doing so is to save five people. Now suppose that you are asked whether it is acceptable to pull a switch that drops someone through a trapdoor, when the result of doing so is also to save five people. Suppose that you believe that it is indeed acceptable. Now suppose that you are asked to square your judgments in the two cases.[76] You might decide that you cannot – that in the first case, physical contact is making all the difference to your moral judgments, but that on reflection, it is irrelevant.[77] If that is your conclusion, you might be moved in a more consequentialist direction.

And in fact, there is evidence to support this view.[78] Consider this case:

A major pharmaceutical company sells a desperately needed cancer drug. It decides to increase its profits by significantly increasing the price of the drug. Is this acceptable?

Many people believe that it is not. Now consider this case:

A major pharmaceutical company sells a desperately needed cancer drug. It decides to sell the right to market the drug to a smaller company; it receives a lot of money for the sale and it knows that the smaller company will significantly increase the price of the drug. Is this acceptable?

Many people believe that it is. In a between-subjects design, people see the case of indirect harm as far more acceptable than that of direct harm. But in a within-subjects design, the difference evaporates. The apparent reason is that when people see the two cases at the same time, they conclude that the proper evaluation of harmful actions should not turn on the direct-indirect distinction. We could easily imagine the same process in the context of other moral dilemmas. If System 1 and emotional processing are leading to a rapid, intuitive conclusion that X is morally abhorrent, or that Y is morally acceptable, an encounter with cases A and B may weaken that conclusion and show that it is based

---

[76] *See* Paxton et al., *supra* note, at 171.

[77] Things may not, of course, be nearly that simple. A decision to push someone may in fact lead to worse results for the individual – certainly emotionally and perhaps legally as well. A decision not to push the person may result in a life with less guilt and trauma. In this respect, the emotional signal is telling us something important. (I am grateful to Tali Sharot for pressing this point.) Note, however, that the guilt and trauma may themselves be a product of the use of a moral heuristic that is misfiring.

[78] *See* Paharia et al., *Dirty Work, Clean Hands: The Moral Psychology of Indirect Agency*, 109 ORG. BEHAVIOR AND HUMAN DECISION PROCESSES 134, 140 (2009). For relevant discussion, see Berker, supra note, at 327-29.

on morally irrelevant factors, or at least factors that people come to see as morally irrelevant after reflection. It is also possible that when people see various problems at the same time, they might conclude that certain factors are morally relevant that they originally thought immaterial.

The same process may occur in straightforwardly legal debates. Suppose that intuition suggests that punitive damages are best understood as a simple retributive response to wrongdoing, and that theories of deterrence seem secondary, unduly complex, or essentially beside the point.[79] If this is so, people's judgments about appropriate punitive damage awards will not be much influenced by the likelihood that the underlying conduct would be discovered and punished – a factor that is critical to the analysis of optimal deterrence.[80] Critical reflection might lead people to focus on the importance of deterrence and to conclude that a factor that they disregarded is in fact highly relevant. Indeed, critical reflection – at least if it is sustained -- might lead people to think that some of their intuitive judgments about fairness are not correct, because they have bad consequences. Such reflection might lead them to change their initial views about policy and law, and even to conclude that those views were rooted in a heuristic.[81]

Even here, however, people's ultimate conclusions are not decisive. Let us stipulate that people might well revisit their intuitions and come to a different (and perhaps more consistently consequentialist) point of view. Suppose that they do. Do we know that they are right? Not necessarily. Recall that for some problems, people's immediate answers are more accurate than those that follow from reflection. With respect to moral questions, the answer would have to depend on the right moral theory. The fact that people have revised their intuitions does not establish that they have moved in the direction of that theory. It is evidence of what people think, under certain conditions, and surely some conditions are more conducive to good thinking than others. But the question remains whether what they think is right or true.

V. Conclusion

We are learning a great deal about the psychology of moral judgment. It is increasingly plausible to think that many deontological intuitions are a product of rapid, automatic, emotional processing, and that these intuitions play a large role in debates over public policy and law.[82] But the relevant evidence is neither necessary nor sufficient

---

[79] *See* Sunstein et al., *supra* note, at 248-49.

[80] *See* A. Mitchell Polinsky and Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 874 (1998).

[81] *See* STEVEN SHAVELL AND LOUIS KAPLOW, FAIRNESS VERSUS WELFARE (2004). Of course a different kind of movement is possible – if, for example, critical reflection leads people to conclude that consequentialist approaches to slavery, torture, rape, and freedom of speech are ultimately inadequate.

[82] Recall that we do not yet know how many deontological intuitions can be understood in this way; consider nonconsequentialist objections to slavery, torture, censorship, and restrictions on religious liberty. Recall too that we lack much cross-cultural evidence.

to justify the conclusion that deontology is a mere heuristic (in the sense of a mental shortcut in the direction of the correct theory). What is required is a moral argument.

There is one qualification to this claim. It is well-established that with respect to factual questions, rapid reactions, stemming from System 1, generally work well, but can produce systematic errors. Deontological intuitions appear to have essentially the same sources as those rapid reactions. That point does not establish error, but it does suggest the possibility that however firmly held, deontological intuitions are providing the motivation for elaborate justifications that would not be offered, or have much appeal, without the voice of Gould's homunculus, jumping up and down and shouting at us. The homunculus might turn out to be right. But it is also possible that we should be listening to other voices.