

This is the more or less final version of an essay that appeared in *What's Next: Dispatches on the Future of Science* edited by Max Brockman and published by Vintage (2009). Pp 104-115.

Fruit Flies of the Moral Mind

Joshua D. Greene

Consider the following moral dilemma: It's wartime. You and your fellow villagers are hiding from nearby enemy soldiers in a basement. Your baby starts to cry, and you cover your baby's mouth to block the sound. If you remove your hand, your baby will cry loudly, and the soldiers will hear. They will find you, your baby, and the others, and they will kill all of you. If you do not remove your hand, your baby will smother to death. Is it morally acceptable to smother your baby to death in order to save yourself and the other villagers?

It's unsettling to think about such questions, but it's also instructive. This dilemma, known as the Crying Baby dilemma, nicely captures the tension between two major schools of moral and political thought. On the one hand, we have the utilitarians, philosophers like Jeremy Bentham and John Stuart Mill. According to them, acting morally is ultimately a matter of producing the best overall consequences, striving for "greater good." On the other hand we have the deontologists, philosophers like Immanuel Kant who think that rights and duties often trump the greater good. In the Crying Baby dilemma, the greater good (at least in terms of the number of lives saved) is served by smothering the baby. But many would say that smothering the baby, in addition to being tragic and difficult to do, would also be morally wrong—a violation of the baby's rights, the parent's duty, or both.

The Crying Baby dilemma is also a window into the organization of the human brain. People often speak of a "moral faculty" or a "moral sense," suggesting that moral judgment is a unified phenomenon, but recent advances in the scientific study of moral judgment paint a very different picture. Moral judgment, it seems, depends on a complex interplay between intuitive emotional responses and more effortful "cognitive" processes. More specifically, it seems that intuitive emotional responses to harmful actions ("Don't smother the baby!") depend on one set of brain systems, while our more controlled, cognitive responses ("Smothering the baby promotes the greater good,") depend on a different set of brain systems. When we puzzle over such moral dilemmas, these neural systems compete, and our all-too-human sense of anguish is the product of that competition. If I'm right, this tension between competing neural systems underlies not only centuries-old disagreements between the likes of Mill and Kant, but also

contemporary tussles over issues such as stem-cell research and the torturing of suspected terrorists.

Let's consider a pair of moral dilemmas that together give rise to the Trolley Problem, a staple of contemporary ethics. The first of these we'll call the Switch dilemma, and it goes like this: A runaway trolley is about to run over and kill five people, but you can save them by hitting a switch that will divert the trolley onto a side track, where it will run over and kill only one person. Is it OK to hit the switch? Here, most people say Yes, consistent with utilitarian philosophy. Next consider the Footbridge dilemma: Here, too, a runaway trolley threatens five people, but this time, instead of standing by a switch, you are standing on a footbridge spanning the tracks, in between the oncoming trolley and the unsuspecting five. Next to you is a large man, and the only way to save the five is to push this large man off the footbridge and into the trolley's path, stopping the trolley but killing your human trolley-stopper in the process. Is it OK to push this man to his death in order to save the five? (I know what you're thinking, and I'll have none of it: No, you can't jump yourself. You're not big enough to stop the trolley. No, you can't shout a warning to the people on the tracks. Yes, your aim will be perfect and the large man will indeed stop the trolley. No, the large man is not Osama bin Laden and the people on the tracks are not your parents, your two children, and your personal trainer. In short, you may not rewrite the question to make it easier.) In response to this case—properly interpreted—most people judge that it would be wrong to sacrifice one life to save five. And here Kant et al. carry the day, as most people place the rights of the man on the footbridge above the greater good.

Why do we go with numbers in the first case but not in the second? Several years ago I had a hunch that the action in the Footbridge dilemma, with the up-close-and-personal pushing, is more emotionally salient than the action in the Switch dilemma, and that this difference in emotional response could explain why we respond so differently to these two cases. My collaborators and I tested this hypothesis by scanning people's brains while they contemplated dilemmas like the Footbridge dilemma, which we called "personal dilemmas," and dilemmas like the Switch dilemma, which we called "impersonal dilemmas." Our hypothesis predicted that the personal dilemmas would elicit increased activity in parts of the brain associated with emotion, while the impersonal dilemmas would elicit increased activity in parts of the brain associated

with more effortful, cognitive processes, such as reasoning. And that's what we found.¹ More specifically, responding to personal dilemmas, such as the Footbridge dilemma, elicited increased activity in the medial prefrontal cortex, along with other brain regions associated with emotion and social thinking. Impersonal dilemmas like the Switch dilemma, by contrast, elicited increased activity in the dorsolateral prefrontal cortex, a classically "cognitive" part of the brain that becomes more active when, for example, you're holding in mind a phone number.

What does this tell us about moral thinking? Here's the idea: In response to both the Switch and Footbridge dilemmas, people engage in utilitarian reasoning: "Five lives at the cost of one? Sounds like a good deal." But in response to the more personal harm proposed in the Footbridge dilemma, there is also a negative emotional response that says, "No! Don't push that man!" and this response tends to dominate the decision. The emotional response in the Switch dilemma is considerably weaker. As a result, utilitarian reasoning dominates the decision and we vote for saving the five. The emotional response that dominates the decision in the Footbridge dilemma depends on neural activity in emotion-related brain regions such as the medial prefrontal cortex, while the more actuarial thinking that dominates the decision in the Switch dilemma depends on neural activity in classically "cognitive" brain regions such as the dorsolateral prefrontal cortex.

In a follow-up experiment, we focused on more difficult dilemmas such as the Crying Baby case. These, too, are personal dilemmas, but they're constructed so that the utilitarian rationale is stronger. In the footbridge case, it's one life versus five, but in the Crying Baby case *everyone* dies if you don't act, including you and your baby. In response to the Switch and Footbridge dilemmas, people's judgments are fairly consistent, but in response to the Crying Baby Dilemma people's judgments are split about 50/50, and nearly everyone takes a long time to respond. What's going on? If the theory I've described is correct, the Crying Baby dilemma triggers a conflict between emotional and "cognitive" parts of the brain. Conveniently, there is a part of the brain called the anterior cingulate cortex that reliably responds to this kind of internal conflict. When your brain is trying to do two different things at once, the anterior cingulate cortex says, "Houston, we have a problem." We predicted that this area would become more active in response to dilemmas like the Crying Baby case, and indeed it does.

¹ J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, & J. D. Cohen, J. D., "An fMRI Investigation of Emotional Engagement in Moral Judgment," (2001) *Science*, 293(5537): 2105-08.

If the anterior cingulate cortex says, “Houston, we have a problem...” this naturally raises the question: Where’s Houston? Houston, it turns out, is in the dorsolateral prefrontal cortex. Once again, this part of the brain enables us to hold phone numbers in mind and engage in abstract reasoning. It also gives us our ability to resist impulses. The common theme among these operations is *cognitive control*—the ability to guide attention, thought, and action in accordance with goals or intentions. Above I said that the thought of harming someone in a “personal” way triggers an emotional response that makes us say, “No!” If that’s right, then approving of a “personal” harm because it will promote the greater good requires the ability to override that emotional response. And *that* requires increased activity in the dorsolateral prefrontal cortex, the seat of cognitive control. This suggests that when people make utilitarian judgments in response to difficult dilemmas like the Crying Baby case, they should exhibit increased activity in their dorsolateral prefrontal cortices, which is what we found.² A more recent study of ours fits the same pattern. We had people consider dilemmas in which promoting the greater good requires breaking a promise, and, as before, we saw more activity in the dorsolateral prefrontal cortex when people gave utilitarian answers favoring the greater good.

This “dual-process” theory of moral judgment—“dual-process” because it posits distinct emotional and cognitive processes—makes some interesting predictions about the behavior of neurological patients. For example, patients with frontotemporal dementia (FTD) are known for their “emotional blunting.” A team from UCLA presented FTD patients with versions of the Switch and Footbridge dilemmas. Their responses to the Switch dilemma were pretty standard, but they were far more likely than others to approve of pushing the man off the footbridge. Without the emotions to tell them “No!” this action, too, seemed like a “good deal.” Two other research teams, one in Iowa and one in Italy, got similar results testing patients with damage to the ventromedial prefrontal cortex, a region known to be important for emotion-based decision making. Both groups found that these patients gave unusually utilitarian responses to dilemmas like the Footbridge and Crying Baby cases. The Iowa patients, in fact, were almost five times more likely than control subjects to give utilitarian responses.³

² J.D. Greene, L. E. Nystrom, A. D. Engell, J. M. Darley, & J. D. Cohen, “The Neural Bases of Cognitive Conflict and Control in Moral Judgment,” (2004) *Neuron*, 44(2),389-400.

³ E. Ciaramelli, M. Muccioli, E. Ladavas, & G. di Pellegrino, “Selective Deficit in Personal Moral Judgment Following Damage to Ventromedial Prefrontal Cortex,” (2007) *Social Cognitive and Affective Neuroscience*, 2(2), 84-92. M. Koenigs, et al. , “Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements,” (2007)

Carlo Valdesolo and David DeSteno at Northeastern University used a clever, low-tech trick to make the same point. They presented people with versions of the Switch and Footbridge dilemmas under two different conditions. At the start of the experiment, some people watched a funny film clip from *Saturday Night Live*, while others watched a clip with no particular emotional content. People's responses to the Switch dilemma were unaffected by the choice of film, but the ones who watched the funny SNL clip were almost four times more likely to approve of pushing the man off of the footbridge.⁴ The idea here is that a dose of positive emotion can neutralize the negative emotion that would otherwise make people uncomfortable with pushing the man off the footbridge.

My colleagues and I conducted a similar experiment, targeting cognitive control processes rather than emotional ones. In our experiment, people had to make their judgments while simultaneously keeping an eye on a stream of numbers scrolling across the computer screen. Every time the number 5 went by, they had to hit a button. This kind of nuisance task is known as a "cognitive load," the purpose of which is to gum up the sorts of higher-level cognitive processes that are based in the dorsolateral prefrontal cortex. We found that the cognitive load made people slower to give utilitarian answers ("Smother the baby in the name of the greater good") but had no effect on the characteristically deontological answers ("Don't smother the baby, even if everyone will die"). (In fact, the cognitive load seemed to *speed up* the deontological answers, but this effect was not statistically significant for the group as a whole.) These two studies are like mirror images: Block the emotional processes, and utilitarian judgments come more easily. Block the controlled cognitive processes, and utilitarian judgments come more slowly.

These results are part of a more general pattern, one that philosophers may find surprising. When an apparent moral duty ("Don't use people as trolley-stoppers") conflicts with the greater good ("Better to save five lives"), judgments in favor of duty are driven by emotion, while judgments in favor of the greater good are driven by more controlled cognitive processes. This is surprising because philosophers like Immanuel Kant, who place duty above the greater good, are often regarded as "rationalists," philosophers whose moral conclusions are supposed to

Nature, 446(7138), 908-11. M. F. Mendez, E. Anderson, & J. S. Shapira, "An Investigation of Moral Judgment in Frontotemporal Dementia," (2005) *Cogn. Behav. Neurol.*, 18(4), 193-97.

⁴ P. Valdesolo & D. DeSteno, "Manipulations of Emotional Context Shape Moral Judgment," (2006) *Psychol. Sci.*, 17(6), 476-77.

be grounded in reason. But the studies described here suggest that this kind of philosophy is less about reasoning and more about *rationalizing*. My colleague Jonathan Haidt thinks that pretty much all of moral reasoning is like this,⁵ but I disagree. Based on the research described above, I believe that utilitarian judgments really are driven by reasoning processes, enabled by the dorsolateral prefrontal cortex, a.k.a. “Houston.” Of course, I don’t think it’s all so neat and simple. Following the philosopher David Hume, I suspect that even utilitarian calculation requires a kind of emotion—less like alarm bells going off and more like sand accumulating on a scale—and there are indeed hints of this in the brain imaging data.

People sometimes ask me why I bother with bizarre hypothetical dilemmas. Shouldn’t we be studying *real* moral decision-making instead? To me, these dilemmas are like a geneticist’s fruit flies. They’re manageable enough to play around with in the lab, but complex enough to capture something interesting about the wider and wilder world outside. With that in mind, let me introduce a final pair of moral dilemmas, originally devised by the utilitarian philosopher Peter Singer.⁶

You’re walking by a pond one day, when you spot a small child drowning in the water. You could easily wade in and save her, but this would ruin your stylish new Italian suit. So you walk on by. Are you a terrible person? Yes, we say. Next case: You receive a letter from a reputable international aid organization such as UNICEF or Oxfam. They would like you to donate \$500, which they will use to save the lives of several poor African children in desperate need of food and medicine. You feel sorry for these children, but you’ve had your eye on a stylish new Italian suit, and you’d prefer to save your money for that. You toss the letter in the trash. Are you a horrible person? You’re no saint, we say, but you certainly haven’t done anything wrong.

What’s the difference between refusing to save a child who’s drowning right in front of you and refusing to save a child who’s drowning in poverty on the other side of the world? Your rationalizing mind is already at work on the problem: In the case of the drowning child, you’re the only one who can help, but many others can help those poor African children. They’re the world’s problem, not yours. Fair enough. But what if you’re standing around the pond watching

⁵ Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.

⁶ P. Singer, “Famine, Affluence and Morality,” (1972) *Philosophy and Public Affairs*, 1: 229-243.

the child drown with dozens of other people, all of whom are rather fond of *their* stylish Italian suits? (It's the annual meeting of the American Bar Association.) *Now* is it OK to let the child drown? We can play this game all day, but we are unlikely to find a satisfying resolution because the human mind wasn't designed to achieve rational moral consistency. An alternative approach is to think about the relevant psychology and its natural history.

Let's try this first with the Switch and Footbridge dilemmas. As explained above, pushing someone to his death is more emotionally salient than hitting a switch that achieves the same thing. But why? An evolutionary perspective may be useful. We evolved in an environment in which good-old-fashioned pushing and shoving were prevalent, but we did not evolve in an age of mechanically mediated threats. It makes sense, then, that these more basic forms of "personal" violence push our moral buttons, while distinctively modern forms of violence do not. Something similar may be true for acts of altruism and the emotions that support them. We did not evolve in an environment in which one could save the lives of distant strangers by being less fashionable, but we did evolve in a world in which one could help desperate people in the here and now. Nature endowed us with tuggable heartstrings, a crucial design feature for creatures whose survival depends on cooperation. But nature couldn't foresee that our survival might someday depend on cooperation across oceans and continents, and so neglected to outfit us with heartstrings that are readily tugged from a distance.

We are, of course, a very clever species. Through our ingenuity, we've made ourselves faster and stronger and more dangerous than all the other creatures on Earth. Perhaps, by applying our capacity for complex cognition to the problems of modern life, we can transcend the limitations of our moral instincts.