Greene et al. (2001) describe a series of experiments designed to gauge the role of emotional engagement in moral decision-making.\(^1\) Their particular focus is narrow: can brain science help solve the puzzle of seemingly inconsistent responses subjects tend to give when confronted with certain emotionally engaging moral thought experiments? More broadly, however, their work raises the possibility of a significant role for cognitive science in the field of descriptive ethics. That in turn points to a question of ecological validity—if cognitive science can enhance our understanding of why people make the moral judgments they do, what bearing would that new understanding have on people who are neither cognitive scientists nor moral philosophers?

Greene et al. note the significance of non-academic implications explicitly, concluding, “The present results raise but do not answer a more general question...How will a better understanding of the mechanisms that give rise to our moral judgments alter our attitudes toward the moral judgments we make?” (p. 2017). Though their results do not entirely answer this question, the question itself demonstrates one kind of ecological validity that could stem from the partnership between cognitive scientists and descriptive ethicists. Understanding how our moral intuitions arise on the algorithmic level—in effect, curing our instinct blindness to the complex cognitive processes that may underlie our notions of right and wrong—would surely change the way we view the reliability of those intuitions as answers to our moral questions at the computational level. Regardless of how our view might change, the very fact that it would change—through self-consciousness of the mechanisms underlying our moral judgments, and the following recognition of the interference that might result from those processes—constitutes a very significant kind of ecological validity.
The example of the Greene et al. study makes this point very clear. The authors ask why subjects will consistently approve of the choice (a) to push a switch and divert an oncoming trolley from killing several people, instead killing only one, but nevertheless disapprove of the choice (b) to push a single person off a footbridge into the path of the trolley, with the same result. Using fMRI data, Greene et al. conclude that “personal” moral dilemmas, characterized by the need to personally interact with the “losers” in one’s proposed resolution of a conflict situation (e.g., pushing a person, rather than pushing an “impersonal” switch with the same result) engage emotional centers of the brain which may account for the seemingly irrational inconsistency in subjects’ responses. One can easily imagine that a person familiar with the findings of this study, who then confronted a dilemma of the “moral-personal” variety in the real world—for instance, the footbridge trolley problem—might be significantly more likely to doubt her own intuitive aversion to pushing the man, realizing that this moral intuition may derive not from some rational process (even one she is unable to articulate), but rather from emotional interference. Recognizing the way our moral decision-making is mediated by our underlying cognitive mechanisms has great ecological importance, then, because such recognition might lead this person to subdue her emotions and push the man over.

Generalizing from this example, cognitive science could find great ecological validity in its capacity to provide essential information for the ethical school of moral liberationism, which holds pure reason supreme and believes that we should overcome (“liberate” ourselves from) our irrational intuitions. Studies like that of Greene et al. could help to inform liberationists of where precisely, on a cognitive level, those irrational obstructions lie. Practical examples go well beyond learning how we ought to act in the unlikely footbridge scenario. For instance, the work of one prominent liberationist, Peter Unger, centers on the moral intuitions of middle-class
Westerners regarding global humanitarian crises. In the remainder of this essay, I will argue that the potential ramifications of experimental cognitive research for Unger’s work alone demonstrate the ecological validity of the ethical domain of cognitive science. Accordingly it will be necessary to outline Unger’s project, which has not yet been significantly informed by cognitive science, in some detail.

Unger sets out to better understand the natural assumption of a dichotomy between doing and allowing harm to others, and the consequent general indifference among Westerners to devastating but preventable illness among children throughout much of Africa and Asia. He proposes a series of thought experiments, gauges subjects’ responses to each, and then tries to account for the differential reactions he observes. Here are the two salient probes, paraphrased briefly:

(1) **Bob’s Bugatti.** Bob is out driving his very expensive Bugatti when he parks it alongside the trolley tracks. He gets out, goes for a walk along the tracks, and notices a runaway trolley barreling down the track. He looks up the track and sees a child trapped in the path of the trolley. He can’t stop the trolley or free the child, but he can flip a switch to divert the trolley down the track he came from, saving the child and destroying his expensive car. He evaluates this, decides the child isn’t really *his* problem, and remembers how valuable his car is to him, so he chooses not to flip the switch. The child dies and Bob drives off.

(2) **John’s Envelope.** John receives an appeal from UNICEF in the mail. After reading it through, he correctly believes that, unless he sends in a check for $100, over thirty children will soon die who would have lived long lives had he sent the money. He throws the material away and does nothing, and the children die.

Unger found that people are intuitively very critical of Bob’s choice, but entirely comfortable with John’s—John could have done something better than what he did, but he did nothing *wrong* in failing to give to charity. This is puzzling for moral philosophers like Unger because, treated rationally, John’s choice was ethically worse than Bob’s. Neither “did” harm, both only allowed it; and whereas Bob’s inaction allowed one death in order to save his
expensive car, John allowed thirty deaths in order to save just $100. Something must account for our irrational intuitions, and Unger suggests the possibility that our moral intuitions are skewed by any of several possible psychological factors.

First, he proposes a sense of urgency or occupying the same “situational territory,” of being involved, that makes us more sympathetic to the child in the trolley problem than in the envelope scenario. This might stem from the physical notion of a switch, from the image of a rushing trolley, or from an inclination to “see” a child on the tracks. Second, he proposes a sense of futility that makes us engage less with the children overseas than with the child on the tracks. In the trolley problem, Bob completely controls the outcome, but in the envelope scenario millions of children will be dying whether John saves thirty or not, and this somehow leads us to feel—but not to rationally believe—that, given the scale of the problem, those thirty lives are insignificant.

This work could dovetail perfectly with experimental cognitive science, in the vein of the Greene et al. fMRI study. The problem Unger raises is clearly one of tremendous real-world significance—if a person would not let a child die in order to preserve a luxury car, why would that same person buy a luxury car to begin with, rather than using that money to save thousands of children? If we could answer this question, locating the source of this irrationality, we could construct policy initiatives, media campaigns, and other interventions that target it for repair. Unger proposes a handful of plausible psychological explanations, but as a conventional descriptive ethicist, he cannot validate them. From the perspective of cognitive science, however, these hypotheses and others seem ripe for empirical testing and subsequent validation or rejection using an fMRI methodology like that employed by Greene et al. A subtraction method study that identified what a brain “feeling involved in the situational territory” of a probe
scenario looks like, compared to a brain “feeling detached from the situational territory” of the probe, could be used to support or reject Unger’s first proposed explanation. Experiments could similarly be constructed to capture the “futility” explanation, and others. Indeed, a reverse-correlation method like that used in the Hasson et al. (2004) study of natural vision could allow investigators to go beyond testing Unger’s proposed explanations, instead generating entirely new hypotheses based on the areas of the brain activated by different moral thought experiments (e.g., Bob’s scenario as compared to John’s).

The point here is that by providing empirical evidence for particular explanations of ecologically important irregularities in moral judgment, the cognitive science of descriptive ethics could have dramatic real-world implications. If, for instance, fMRI studies found evidence of greater activation in a visual center of the brain when subjects are confronted with the trolley problem than with the envelope problem, because (hypothetically) we are more inclined to visualize dramatic or exciting scenarios than comparatively mundane ones, we might conclude (after controlling for confounding explanations) that this visualization of the child on the tracks increases our sympathy for him. That in turn might lead international charities to develop an advertising message employing more visually dramatic stimuli. Similarly, if cognitive studies found a greater recognition of time-criticality in the brains of subjects confronted with the trolley problem than those confronted with the envelope problem, that might suggest a campaign emphasizing the rate at which children are dying, conveying to the audience that they, like Bob, must act immediately if they are to save the children in question. But the greatest and most unique power of ethical cognitive science might lie in reverse-correlation studies that would highlight factors or explanations that could not be listed here, because we are not yet even aware of their potential importance.
Two objections could be made to the argument for ecological validity presented here. First, an opponent of this view could rightly point out that the analysis above is predicated on the assumption that moral reasoning ought to be a rational deductive process, in which we apply general principles (e.g. saving $x$ at the cost of $y$ is good, so long as $x > y$) to specific cases (e.g., saving five people at the cost of one is good) and infer a course of action (e.g., push the man off the footbridge). If this kind of rationality is accepted as the proper basis for moral judgment, then as I have argued, cognitive science offers a powerful opportunity to liberate ourselves from the psychological mediators that might interfere with our moral deductions, just as consciousness of our own loss-aversion improves our non-moral decision-making. But if, on the other hand, one rejects the basic assumption that rationality should underlie morality, one would likely also reject this argument for the ecological validity of a cognitive approach to descriptive ethics. That is, if our irrational intuitions are treated not as interfering with right moral decision-making, but rather as a valid basis for moral judgment, then it would no longer be to our advantage to examine the psychological roots of those intuitions so as to override them with reason. Two responses to this objection present themselves. First, this favorable view of irrationality is not shared by the mainstream of our ethical tradition, which is deeply rationalist. Aristotle, Kant, and Mill all derive humanity’s unique capacity for moral judgment from our unique faculty of rational thought—and they all share the view that our visceral responses do not constitute moral justifications.

But furthermore, even if we do not seek to make our moral judgment algorithms transparent for the purpose of subjugating them to pure reason, there are other ecologically valid motivations for understanding the lower-level causes of our moral intuitions. Even those who propose a morality ungrounded in reason tend to view some irrational ethics as superior to
others. Within a system of rationality, one can understand, evaluate, and criticize others’ ethics for being irrational. But for these irrational ethicists, there is no such universal standard. Thus it is precisely the irrationalists who might reap the most profound benefits from a cognitive investigation of the algorithms that produce our ethical intuitions. Cognitive scientists could investigate whether a propensity for “bad” ethical intuitions is innate, and whether this results from a malfunction or malformation of the brain’s ethical hardware (e.g., a short-circuit that bypasses the emotional center highlighted by Greene et al.). If this were the case, the irrationalists would not only be able to explain ethical deviance, but might be able to develop interventions that target it for therapy at the level of the underlying mechanism—albeit in a fashion reminiscent of *A Clockwork Orange*.

A second noteworthy objection to the ecological validity of this domain of cognitive science would grant a rationalist morality, but question its significance in the real world. That is, even if this research could subject our opaque and possibly error-prone ethical judgment systems to rational scrutiny, and thus inform people that John’s envelope scenario is “actually” no different from Bob’s trolley scenario, would anyone really act differently because of this new knowledge? This is effectively the problem Aristotle describes as “incontinence”—knowing what is right, but still lacking the fiber to do it. This objection can be answered in two ways as well.

First, realizing that we confront two distinct problems in encouraging ethical action—one of determining the right act, and the other of developing the resolve to do what we believe is right—there is no reason to assume that cognitive science could only tackle the former problem. Indeed, experiments could be conducted in which participants not only pass judgments on moral scenarios while under fMRI observation, but also actually do (or fail to do) moral acts from
within the fMRI scanner (e.g., transfer money from their bank accounts to charities). Secondly, even if a cognitive investigation of ethics sheds no light on the problem of moral continence, attacking the problem of moral judgment alone would still matter in the real world—first because there are many continent individuals who nevertheless act irrationally, and secondly because progress on one front would allow conventional ethicists and policymakers to devote more resources to the other.

In sum, the cognitive science of descriptive ethics matters practically on both the individual and the societal levels. By helping us to identify the specific psychological mediators that yield irrational moral judgments, this research will enable us to better master those sometimes-irrational intuitions, just as readers of the Greene et al. study would be more likely to push the man from the footbridge. At the same time, this research will help us to design advocacy campaigns that frame unexciting but crucial moral problems—like millions of inexpensively preventable deaths each year—in terms that appeal to people’s moral sensibilities, emphasizing in the real world those dimensions of a problem that have proved most salient in the laboratory.

References


