



Some Socratic Modesty: A Reconsideration of Recent Empirical Work on Moral Judgment

David Sackris¹ · Michael T. Dale²

Accepted: 15 April 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Plato's *Euthyphro* is famous for the dilemma Socrates presents those who subscribe to divine command theory. Yet, it's worth remembering what prompts the discussion: Socrates' incredulity that Euthyphro would prosecute his own father for murder (4a-b, trans. Grube). Socrates is shocked that someone could be so confident in their moral judgment, especially as he learns more about the case: The person whom Euthyphro's father killed also killed someone. One way, then, to understand the dialogue, is like this: Socrates aims to show that Euthyphro does not have any special moral knowledge, and that it's harder to obtain moral knowledge than we might think.

One way to interpret the work of Joshua Greene (2001; 2008; 2014) is that the recent wave of empirical research into moral decision-making is a way for us to become *more* confident in our ability to gain moral knowledge.¹ If we can understand the cognitive bases of some of our moral judgments, we might not only learn how we do make moral judgments, but also how we should make those judgments.

Here we aim to strike a more skeptical tone. We shall argue that research into moral judgment should make us more skeptical about our ability to resolve moral dilemmas, not less. We review empirical research from the past twenty years into moral judgment (both survey-based and brain-based) that has shown us that the grounds of moral judgment are opaque. We largely agree with these findings, but we also aim to generate a deeper sort of skepticism. To do so, we advance a novel criticism of Greene's argument that aims to show not just that study participants base their responses to moral dilemmas on seemingly irrelevant factors, but that researchers are themselves unclear on what factors moral judgments should be responsive to.

¹ Singer (2005) endorses this interpretation.

✉ David Sackris
david.sackris@gmail.com

Michael T. Dale
mdale@hsc.edu

¹ Arapahoe Community College, 5900 S Santa Fe Drive, Littleton, CO 80120, USA

² Hampden-Sydney College, 172 Via Sacra, Hampden Sydney, VA 23943, USA

² We show that factors about the decision context and the decision-maker play a role in reaching moral verdicts in ways that demonstrate that we lack insight into both how we do make moral judgments and how we should make moral judgments.³

We begin by reviewing research into moral judgment via the use of trolley problems. We then consider Greene's attempt to resolve the trolley problem and its iterations through his employment of Dual Process Theory (DPT). We argue that even if DPT is true, it is not clear that it can be used to help us gain moral knowledge because we have no clear grounds for preferring one cognitive system over another. Our argument for this makes use of a position that is rapidly gaining acceptance in the field of moral psychology: Moral judgments are not different in kind from our general judgments. Research in a variety of fields, such as the field of aesthetics and economics, shows judgments are highly influenced by the judgment context. If moral judgment is just another species of judgment, as many researchers (including Greene) have come to believe, we shouldn't be surprised to learn that moral judgment is often based on seemingly arbitrary factors. We conclude by arguing that empirical research into moral judgment should encourage us to have an ironic attitude (in the Rortian sense)⁴ towards our moral judgments: We cannot help but make them, but we should keep in mind that they are highly influenced by contextual and individual factors out of our control.

1 Trolley Problems, Dual Process Theory, and Moral Knowledge

The contemporary study of moral decision-making owes much to Judith Thomson's (1976) and Phillipa Foot's (1967) discussion of the so-called "Trolley Problem". In her exploration of the distinction between killing and letting die, Thomson gifted the philosophical community the following cases:

Switch: An individual can throw a switch to divert an out-of-control trolley from killing five but in the process leads to the death of one individual on a sidetrack.

Footbridge: An individual can stop a trolley from killing five but only by pushing a very heavy individual onto the track, thereby killing that individual.⁵

² Berker (2009) argues that Greene needs to know what features of a dilemma situation are morally relevant, but he does not make a skeptical argument regarding our ability to come to know these things. For criticisms of Greene, see also Bruni, Mameli, and Rini (2014); Kamm (2009); Königs (2018); Kumar and Campbell (2012).

³ Others have argued that the basis for moral judgment is influenced by seeming irrelevant contextual factors (e.g., Haidt 2012; more recently Sauer 2021; Klenck and Sauer 2021), but they have not argued that we are unclear on which factors we *should* base our decisions. Further, Sauer (2021) reaches a much more radical conclusion that we do here. We will explain why we do not embrace his "moral abolitionism" subsequently.

⁴ See his 1989.

⁵ These are not the exact cases they employ, but the set up comes from the debate between Thomson and Foot.

Thomson states “presumably” the bystander “may not shove the fat man into the path of the trolley” (1976, p. 208); however, she thinks the bystander *should* divert the train by pulling a switch so that it kills one instead of five in *Switch*. When lay-people have been presented these problems, they have tended to share her intuitions (Bauman et al. 2014; Cushman et al. 2006; Greene et al. 2001; Petrinovich et al. 1993); that is, they approve of killing one to save five in *Switch* but not in *Footbridge*.⁶ In a 2014 study conducted by David Bourget and David Chalmers, 68.2% of professional philosophers indicated they would divert the trolley in *Switch*, thereby killing one to save five; 7.6% said they would not divert the trolley, and 24.2% choose “other” (2014, p. 16). In their 2023 follow up study, they found that only 22% of professional philosophers were willing to push the large man off the foot-bridge (p. 6).

That people do treat *Switch* and *Footbridge* differently does not mean that they should. That is, why should having to lay hands on a person make a moral difference to what we should do? Are we too eager to pull the switch in the *Switch* or too reticent to push the large man in *Footbridge*? Resolving this dilemma seems to require certain moral knowledge: It requires us to know which features of the situation are morally relevant. For example, thinking that pushing the heavy person in *Footbridge* is the morally right thing to do would require us to have good reasons for rejecting the doctrine of double effect.⁷

Greene doesn't see things this way. He contends that we have good reasons for thinking that diverting the trolley in *Switch* and pushing the large man in *Footbridge* are the morally right actions. These reasons are put forward in two seemingly separate (but interrelated) arguments (Greene, 2008; 2014). The first argument claims that the decision not to sacrifice one person to save five in *Footbridge*—i.e., the characteristically deontological judgment—is guided by an emotional reaction to the morally irrelevant factor of up close and personal harm. Because people emotionally reject laying hands on another person to cause harm, they reject acting in *Footbridge*. However, there is no strong emotional aversion to acting in *Switch*. Because Greene takes the proximity of the harm to be morally irrelevant, we have reason to doubt the reliability of the judgment in *Footbridge*.⁸ Greene is confident that those who reject acting in *Footbridge* do so because of a factor that should not be influencing their judgment.

Greene's second normative argument stems from his subscription to DPT. DPT maintains that there are two distinct ways in which the human mind problem solves, and these two approaches are typically referred to as System 1 and System 2.⁹

⁶ Although Greene et al. 2009 does not directly compare *Switch* and *Footbridge*, they find that subjects generally rate the permissibility of pushing one to save five lower than other sorts of actions that involve killing one to save five.

⁷ The doctrine of double effect holds that it may be permissible to cause someone harm to bring about some greater good if the harm is a side-effect of bringing about the good in question, but not if the harm is a means of bringing about the good in question (See McIntyre 2023). Greene (2014) discusses the doctrine of double effect and argues that it is a mere codification of unjustified System 1 intuitions.

⁸ For a formalization of this argument, see Paulo, 2019: 7-10; Dale & Gawronski, 2022: 4.

⁹ For an overview of the development of dual-process accounts of human cognition, see Evans (2012).

System 1 processes are characterized as unconscious, rapid, low effort, intuitive, associative, emotion-based, and highly contextual; System 2 processes are described as conscious, controlled, slow, engaged in abstract reasoning, and as more dependent on individual intelligence (Kahneman 2011; Evans 2012; Mugg 2016, 2018; Greene et al. 2008; Greene 2014). For example, in some of the original work on judgment that gave rise to DPT, when subjects give incorrect answers to logic puzzles or misunderstand the importance of sample size to probabilistic outcomes, this is attributed to the operations of System 1: The subject went with the answer that most quickly came to mind instead of more carefully analyzing the question.

Greene believes that when individuals decide to sacrifice one person to save five in both *Switch* and *Footbridge*, these verdicts are delivered by System 2 processes. Furthermore, we have good reason to prefer System 2 processes when it comes to solving unfamiliar* problems (2014, p. 714). By “unfamiliar*” Greene means “problems with which we have inadequate evolutionary, cultural, or personal experience” (2014, p. 714). According to Greene, System 1 is good for problems with which we are familiar*, but solving certain kinds of novel (i.e., unfamiliar*) problems correctly requires effortful cognition. Greene believes that *Switch* and its ilk constitute unfamiliar* problems. He offers us the following general principle for guiding moral decision making:

No Cognitive Miracles Principle: When we are dealing with unfamiliar* moral problems, we ought to rely less on automatic settings (automatic emotional responses [System 1]) and more on manual mode (conscious, controlled reasoning [System 2]), lest we bank on cognitive miracles. (2014, p. 715)

By “cognitive miracles”, Greene means something like this: Being able to successfully carry out what would otherwise be a cognitively demanding task with no prior training. He says, “If one could drive like an experienced driver from the outset, that would be a cognitive miracle” (2014, p. 714). Typically, when a person is sufficiently trained on some activity, they can successfully carry it out using System 1.

For Greene, we have good reason to tend towards utilitarian judgments when confronted with unfamiliar* moral problems, as utilitarian judgments tend to be rooted in brain areas thought to be responsible for more cognitive processes (i.e., System 2); however, when subjects judge acts in a more deontological fashion, centers of the brain believed to be responsible for emotion are more likely to be active (i.e., System 1) (Green 2001; Greene et al. 2004; Green et al., 2008). Essentially, Greene is identifying deontological-type judgments with knee-jerk emotional reactions and utilitarian-type judgments with controlled cognition. Put another way, Greene believes that System 1 judgments (the system Greene identifies with deontological judgments) are more likely to be influenced by irrelevant biases in unfamiliar* decision contexts.

Presumably, people have not thought much about the strange variations of the trolley problem dreamed up by philosophers and then presented to study participants for judgment, which gives us reason to think they are unfamiliar*.¹⁰ Furthermore,

¹⁰ This will be challenged subsequently.

that there is disagreement about what to do indicates that individuals may have clashing intuitions about such cases, which Greene construes as evidence in favor of a theory on which distinct cognitive systems offer conflicting verdicts (2014, p. 716). In such unfamiliar* cases, we should distrust our System 1 intuitions and go with the product of more effortful cognition. Greene believes that it is System 2 processes that indicate to subjects that they should push the large man off the bridge.¹¹

Although Greene is often interpreted as giving two different arguments in favor of utilitarian judgments in *Switch* and *Footbridge*, we see these as really two argumentative sides of the same coin. System 1 and System 2 judgments map directly onto what Greene earlier referred to as deontological-style and utilitarian-style judgments. It is fast, emotion-based (System 1) judgments that correspond to deontological-style judgments, and cognitively-controlled (System 2) judgments that correspond to utilitarian-style judgments. Furthermore, it is deontological-style System 1 judgments that are supposedly more prone to irrelevant biases on Greene's original (2008) telling of the story.

It is not just on trolley problems where Greene believes we tend to wrongly rest on our System 1 intuitive judgments: In his (2004) he presents individuals with *Crying Baby*. In this dilemma a subject must decide whether it would be appropriate for a parent to smother their baby in order to save themselves and many others from an enemy that plans to kill everyone they find (p. 390). Concerning this dilemma, Greene et al., say:

According to our theory, this dilemma is difficult because the negative social-emotional response associated with killing one's own child competes with a more abstract, "cognitive" understanding that, in terms of lives saved/lost, one has nothing to lose (relative to the alternative) and much to gain by carrying out this horrific act (2004, p. 390).

Although Green doesn't say this in his (2004), his (2014) makes clear what he takes himself to know: Under the assumption that *Crying Baby* is unfamiliar*, the correct thing to do is smother the baby. Such a verdict is delivered by a System 2 cost-benefit analysis, while our unwillingness to smother even one single baby for the significantly greater good is the result of a System 1 emotional aversion. That is, what Greene referred in 2004 to as "negative social-emotional response" is now being identified with System 1, and "'cognitive' understanding" is being identified with System 2.

Other moral dilemmas that we might resolve by subscribing to Greene's normative conclusion include: "climate change, global terrorism, global poverty, bioethics" (2014, p. 716). If Greene is right, there is hope that the solution to a great number of moral problems is in the offing.

¹¹ For an overview of this evidence, see his 2014, especially pp. 701-705.

2 Some Socratic Modesty

We might ask Greene and others the very same question Socrates asked Euthyphro: How can you be sure, even with the empirical evidence you appeal to, about what the morally right thing to do is in these dilemmas? Because even if we can be sure which processes are resulting in which verdicts (which we will call into doubt), it's still not clear how we get to the point at which we *know* which verdict is correct, for there's much else we need to know to reach that conclusion with confidence.

When we consider the trolley problem as a kind of paradox, a key element of the paradox is that throwing the switch in *Switch* is intuitively morally right (or we at least know it's morally permissible) and that pushing the heavy man feels wrong, otherwise there would be no trolley problem (Königs 2023). It is only if we *feel confident* that throwing the switch in *Switch* is morally correct does it matter if *Switch* and *Footbridge* are morally analogous. A number of thinkers on this issue have taken it as obvious that the switch should be thrown even before Greene utilized DPT and fMRI research to support that conclusion (Fischer and Ravizza 1992; Kamm 2007; Singer 2005; Thomson 1976;).¹² For Greene, a key element of 'solving' the trolley problem is the idea that, because of its impersonal nature, people are deciding strictly based on the numbers in *Switch*, which is what he believes they should do.¹³ Therefore, if individuals are reluctant to act in *Footbridge* even though it's clear to Greene the cases are analogous, it must be that some irrelevant factor is leading them astray.

Nonetheless, persistent doubt remains among a substantial minority of philosophers and laypeople, as demonstrated by the survey data. Are we certain that throwing the switch in *Switch* is the right thing to do? And more importantly for Greene, are we sure that the judgment to throw the switch is delivered by System 2? Without establishing these two claims, his argument fails.

Our anecdotal experiences in the classroom are in line with the usual survey results: Two thirds of students endorse throwing the switch in *Switch* and the remaining third is typically divided between rejecting throwing the switch and refusing to take a position. However, we have also observed that students can be led to change positions on *Switch* through further deliberation and discussion. After students read John Taurek's (1977) "Should the Numbers Count?", they become less sure about what the right thing to do is. Taurek's aim is to get the reader to reconsider what is supposedly the self-evident conclusion in *Switch*. Taurek gives the following argument: Suppose it's the person on the side-track who has the power to throw the switch in *Switch*, thereby diverting the train from killing five and in the

¹² Thomson (2008) later changes her position. See Königs (2023) for a discussion of what makes the trolley problem a problem and an overview of the literature.

¹³ Greene makes several remarks in his unpublished response to Berker that demonstrate his confidence in relying on a utilitarian calculus. E.g., "the perfectly natural, untutored judgment that it's acceptable to turn the trolley in [*Switch*] is *not* an intuitive judgment. It is, I claim, a very simple *reasoned* judgment, the result of explicitly applying a utilitarian decision rule" (p. 18; Greene's emphasis); and "All of the factors that push us away from consequentialism will, once brought into the light, turn out to be things that we will all regard as morally irrelevant" (p. 20).

process killing themselves.¹⁴ Taurek asks whether we would try to persuade this one person on the side-track to sacrifice themselves so that five strangers may live. Taurek states that even making such an argument would feel ridiculous. He imagines how this one person might respond:

Isn't he likely to ask: "Worse for whom?... It is a far worse thing for me that I should die than that they should.... Indeed I wouldn't ask, nor would I expect, any one of them to give up his life so that I, a perfect stranger, might continue to live mine. But why should you, or any one of them, expect me to give up my life so that each of them might continue to live his?" (1977, p. 299).

Taurek states that in not sacrificing one's own life to save five others, the person in question does not do anything morally wrong (p. 300). Taurek is essentially asking: If it is acceptable for the person tied to the track to prefer their own life to five others, why should a third party with the power to choose between the one and the five be morally required to save five?

Our students have found this argument quite compelling. These same considerations led Thomson to revise her own position: "[S]ince [the bystander] wouldn't himself pay the cost of his good deed if he could pay it, there is no way in which he can decently regard himself as entitled to make someone else pay it" (2008, p. 366).¹⁵ Research conducted by Ezio Di Nucci supports our anecdotal classroom observations (as well as Thomson's revised position). In his study, participants considered *Switch* and the following trilemma:

Trilemma: You are standing near the railroad tracks and notice an empty boxcar coming down the tracks, moving fast enough to kill anyone that it hits. If you do nothing, the boxcar will continue along the main track, killing five people who are walking down the main track. There is a switch nearby that you can use to divert the boxcar onto either of two side-tracks that split off from the main track in opposite directions. There is one person walking along the right-side track. So, if you flip the switch to the right, the boxcar will hit and kill this person. Your foot is stuck in the track on the left-side track. So if you flip the switch to the left, you will be hit and killed by the trolley yourself. What should you do?

¹⁴ Taurek's example involves a necessary drug that one person needs all of but that could be split into five parts to save five people instead.

¹⁵ Although Thomson cites Taurek, she doesn't credit him with asking us to consider what someone affected by the moral dilemma would be morally required to do. She draws the same conclusion he does 30 years later.

When study participants were presented with *Trilemma* and then *Switch*, the standard results were flipped: Over 60% of participants were now *unwilling* to flip the switch and divert the trolley. Considering whether they would be willing to sacrifice themselves to save five had an impact on their willingness to sacrifice a stranger to save five (Di Nucci 2013, p. 668).

The point here is not whether Thomson, Taurek, and Di Nucci's study participants are ultimately right. But they make a compelling argument that depends on pointing out a seemingly relevant moral feature that has mostly been overlooked in the literature: We tend to think about moral problems differently from a third-party perspective than from the perspective of someone who is intimately involved in the situation, and perhaps that difference is morally unjustifiable. Our confidence that we should sacrifice one to save five in *Switch*, regardless of which cognitive system is yielding said verdict, should be shaken unless we have an a priori basis for knowing which features of the situation are morally relevant. More significantly, if we aren't sure about what to do in *Switch*, there is no clear basis for generalizing to other cases regardless of whether they're truly analogous.

This data challenges Greene's argument in another, more significant way, and it has to do with how we should interpret this change of mind in our students, Thomson, and Di Nucci's study participants. Here is one interpretation of what happened that should cause concern for Greene's argument: Deeper, conscious reflection on the dilemma prompted by a consideration of previously overlooked features has caused individuals to think that they shouldn't throw the switch in *Switch*. In other words, *System 2* may be indicating that the switch should not be thrown, which is directly contrary to Greene's thesis.

There is an alternative interpretation of the evidence available that is just as troubling for Greene's position. Perhaps what Taurek, Thomson, and Di Nucci have shown is that the intuition that we should throw the switch in *Switch* is the result of System 1 reasoning and always has been. There are two main kinds of evidence that Greene appeals to support his conclusion that throwing the switch is the result of System 2 processing in *Switch*: (a) more 'cognitive' areas of the brain appear to be active when individuals consider impersonal dilemmas (i.e., *Switch*) as compared to more personal ones (i.e., *Footbridge*) (2001; 2014); (b) individuals are conflicted between System 1 and System 2 verdicts when they consider more personal dilemmas (2008; 2009; 2014).¹⁶ However, the evidence appealed to in (a) and (b) doesn't actually prove that judging that one should pull the switch stems from System 1.

Our aim here is not to question Greene's methods, but instead point out that his conclusions are primarily based on the consideration of personal dilemmas, such as *Footbridge*, and not the more impersonal *Switch* (e.g., Greene et al. 2009). Greene would have to admit that employing areas of the brain that have a greater association

¹⁶ By 'impersonal' dilemmas Greene (2014) means dilemmas in which the person deciding what to do does not have any physical contact with their potential victim. In a 'personal' dilemma, the person acting does have physical contact with their potential victim.

with cognitive processing does not, *by itself*, prove that System 2 is active.¹⁷ He doesn't explicitly acknowledge this in his published work, but his example indicates his awareness of this possibility. He says that *learning* to drive a car requires System 2, but once we master how to drive a car, we can do so almost entirely via System 1 (2014, p. 714). This is how we can, e.g., arrive at a familiar destination without having consciously thought about the steps required to drive there.

Although Greene presents System 1 as a primarily emotion-based decision-making process, this is a misrepresentation of the original biases and heuristics literature that DPT is based on. If it were possible to scan an experienced driver's brain while they drive, we would expect to see the more cognitive areas of the brain active the whole time even though such processes could be attributed to System 1. Anything we have a great deal of experience with we can do using System 1, and when we are well trained on some activity we are often better off trusting our System 1 processes. For example, we don't have to, and shouldn't, attempt to consciously calculate distance before we throw a ball, nor when we aim to jump over a small puddle; these are examples of typical System 1 judgments that do not involve emotion. Lack of a clear activation in areas of the brain associated with emotion in an fMRI, by itself, doesn't tell us which system (if there are two systems) is doing the work.¹⁸

In Daniel Kahneman's explanation of Systems 1 and 2, he frequently employs an analogy with chess to demonstrate the machinations of System 1; unlike a novice, a master chess player can simply 'see' what the next move should be without having to rely on effortful cognitive processing:

If an individual has relevant expertise, she will recognize the situation, and the intuitive solution that comes to her mind is likely to be correct. This is what happens when a chess master looks at a complex position: the few moves that immediately occur to him are all strong. When the question is difficult and a skilled solution is not immediately available, intuition still has a shot: an answer may come to mind quickly—but it is not the answer to the original question. This is the essence of intuitive heuristics: when faced with a difficult question, we often answer an easier one instead, usually without noticing the substitution (2011, pp. 12-13).

Based on Kahneman's presentation of DPT and Greene's own example of driving a car, it's entirely possible that subjects are judging *Switch* via System 1 processes: They have significant experience comparing small sums, and they may be substituting a hard question for an easier one: It's difficult to weigh the value of human lives against each other; it's much easier to compare sums. So, they compare sums and arrive at the intuitive conclusion that the switch should be pulled.

¹⁷ Whether most fMRI studies conclusively demonstrate that specific brain areas can be associated with specific cognitive processes is open to significant doubt. See for example Carp 2012; Marek et al. 2022. fMRI-based findings have been debunked in the field of psychopathy research. See for example Deming et al. 2022.

¹⁸ See Mugg (2016) for criticisms of DPT.

Paul Bloom (2011) makes a somewhat similar point by wondering if study participants even think of *Switch* as a moral problem. The scenario involves nameless, faceless hypothetical individuals that the study participant has no connection with. Bloom argues that we could interpret the results of studies of *Switch* in two ways: On Greene's preferred interpretation, in the absence of "emotional distractions" individuals tend to make calculated, utilitarian-style judgments; however, another interpretation seems equally plausible according to Bloom. In the absence of any identifying information regarding the potential victims, we must consider the "possibility that our intuition here isn't moral at all... we treat the dilemma as little more than a math problem (2011, p. 40).¹⁹ Although Bloom doesn't attribute the solution of this math problem to System 1 processes, doing so fits with our analysis. However, if it really is System 1 judging that we should kill one to save five in *Switch*, on Greene's own theory we have good reason to doubt that killing one to save five is the right thing to do since he claims that System 1 is less apt for solving unfamiliar* problems.

If we weren't sure what the solution to *Switch* was before the empirical research program into trolley dilemmas was initiated, we should be less sure now. Taurek, Thomson, and Di Nucci have given us good reason to wonder if we should take into account the perspective of the person who would be sacrificed to save five, and Bloom points out that we cannot be sure what role anonymity is playing. It could be that the anonymity of the victims is causing us to give a simplistic, System 1 analysis. If we consider the 6 people involved as individuals like ourselves who each value their own lives, Di Nucci's data shows we are inclined to give a different verdict. To us, this signals that we cannot be sure what the morally relevant factors of *Switch* are, nor even which cognitive system is taking which factors into account (is it System 1 or System 2 that is merely comparing sums?). So, it looks like we don't know what the morally relevant features of *Switch* are *and* we don't know which system is yielding which verdict. The empirical data should drive us backwards here, not forward.

3 Relevance, Irrelevance, and Footbridge

Let's now turn to *Footbridge*. A closer examination of *Footbridge* is useful for two reasons: (1) adherents of DPT take it to exemplify the conflict between the two cognitive systems; (2) many philosophers have confidently made claims about what is and what is not morally relevant to deciding *Footbridge*.

Greene takes difficult personal dilemmas like *Footbridge* and *Crying Baby* to demonstrate *The Central Tension Principle*:

¹⁹ Whether there is a real difference between moral problems and non-moral problems is irrelevant for the point being made here. If thinking of the participants as identifiable individuals would make participants judge the situation differently, then this is all that matters for the purposes of Bloom's argument. Greene seems to think there is no difference between moral judgments and other judgments, which will be discussed. For arguments that there is no meaningful difference between moral and non-moral judgments, see Sackris (2023); Sackris and Larsen (2022, 2023).

The Central Tension Principle: Characteristically deontological judgments are preferentially supported by automatic emotional responses [System 1] while characteristically consequentialist judgments are preferentially supported by conscious reasoning and allied processes of cognitive control [System 2]. (2014, p. 699)

This distinction is important because, according to Greene's argument, System 1's automatic emotional responses are more likely to be influenced by morally irrelevant factors. In *Footbridge* the initial analysis is that we have a strong emotional aversion to laying our hands on another person, which supposedly competes with our intellectual realization that we should save the greater number of lives. Because the emotional aversion is so strong, most people respond that one should not push the heavy person off the bridge to save five although they might also say that one should divert the trolley in *Switch* (Cushman et al. 2006; Green et al., 2004; Cushman and Greene 2012; Greene, et al. 2009; Greene 2014).

Greene believes we are inappropriately attaching moral relevance to the act of using physical force on someone (2014, p. 723). As we pointed out above, Greene takes this intuition regarding personal force to be "substantive" and "one that nearly all of us share". He says this intuition has the same, obvious status as the following intuition: "Capital juries ought to regard a defendant's race as irrelevant" (ms., pp. 13-14).²⁰ Peter Singer takes a similar, although stronger, position to Greene:

The death of one person is a lesser tragedy than the death of five people. That reasoning leads us to throw the switch in the standard trolley case, and it should lead us to push the stranger in the footbridge, *for there are no morally relevant differences between the two situations....* (2005, p. 350, our emphasis)

Here Singer takes himself to know that there are no morally relevant differences between *Switch* and *Footbridge*, even though a great deal of ink has been spilled on discovering just what the moral differences might be (e.g., Fischer and Ravizza 1992; Gorr 1990; Kamm 2007, 2015; Königs 2023; Thomson 1976, 2008).

In this section we demonstrate that we cannot confidently determine what our moral judgment is responsive to in *Footbridge*, so even if Greene and Singer are right that personal force is a morally irrelevant consideration, it is not clear that this gets us anywhere, as the empirical evidence indicates that our judgment might not be turning on that factor anyway.

Greene's own research lends some support to Bloom's hypothesis regarding the roles of anonymity and its converse, identifiability, in moral decision making. Initially, researchers hypothesized that people treat *Switch* and *Footbridge* differently because of a focus on intention: in *Footbridge*, the large man is used as a means to saving five, but in *Switch* the death of one is the unintended side-effect of saving five (Cushman et al. 2006; Hauser et al. 2007; Schaich Borg et al. 2006).²¹ However,

²⁰ First, it is important to note here that we do think capital juries should disregard the defendant's race. Second, a brief peak at the history of the United States indicates this is not an intuition that has always been universally shared, or taken to be "intuitively obvious" by all jurors at all times.

²¹ That is, they appealed to the doctrine of double effect.

Cushman et al. (2006) also hypothesized that the reluctance to act in *Footbridge* has something to do with physical contact: Having to physically act on another person is what discourages people from acting. Greene also hypothesized that resistance to acting in *Footbridge* has to do with direct physical contact (2001), yet Greene et al. (2009) later found that people were equally resistant to pushing the large person off the bridge with their bare hands as they were to pushing the person with a long pole. So, it is not only physical contact that is deterring people from acting in *Footbridge*—perhaps what is deterring people is that in both the pole case and the pushing case they would be able to ‘identify’ or ‘see the face’ of the person they are killing.²²

However, confounding Bloom’s hypothesis is a further scenario where subjects are asked to imagine that they are standing right next to the large person and they can drop them on the track by pressing a button. In this scenario, subjects were almost as willing to press the button as throw the switch in *Switch* (Greene et al. 2009). As an additional confound to both Greene and Bloom, subjects were willing to use personal force when the person sacrificed to save five is killed as a side-effect of their direct, physical contact (unintentionally knocking a person off a bridge as they rush by them to get to a switch to divert the trolley) instead of as a means to save five (Greene et al. 2009).

In short, to interpret Greene, or anyone else, as having pinpointed what it is that drives decision-making in sacrificial dilemmas like *Footbridge* or its variants would be extremely charitable. At best we can say that in some cases it *seems* like whether the action is intentional or not is driving the judgment of permissibility, and in other cases it *seems* like the exercise of personal force/physically touching the victim is driving judgments of permissibility. What this research really illustrates is that individual’s judgments may be sensitive to a variety of interwoven factors, and that there may be other (possibly morally relevant) factors influencing judgments that we have yet to identify.²³

Now Greene has a ready answer to some of the arguments posed here: Even if people’s aversion to acting in *Footbridge* is due to some mixture of identifiability, physical contact, intentionality, and proximity, these are all likely System 1 intuitions that should be discounted in favor of the cognitive realization that killing someone is killing someone. Just so long as in these scenarios System 2 is in fact indicating that we should sacrifice one to save five, then we have good reason for thinking that sacrificing one to save five is the right thing to do.

However, to discount these as irrelevant emotional aversions, we would have to antecedently know they are morally irrelevant and that they really are the result of System 1 processes, and as argued above it just isn’t clear that we know that. Furthermore, we would have to know that it is just these factors influencing our judgment and not others.

²² See Small and Lowenstein (2003) for evidence that even a very weak form of identifiability increases caring among study participants.

²³ See Berker (2009) for additional criticism of Greene in regards to his ability to precisely pin down what it is that people are responding to in these dilemma scenarios.

The claim that we cannot be sure what factors are influencing our judgment is supported by a recent paper by Paul Rehren and Walter Sinnott-Armstrong.²⁴ There they sought to determine the stability of moral judgments. That is, if you present research participants with standard moral dilemmas, and then re-present them a week later to the same participants, will they give consistent responses? The results are a mixed bag. For some dilemmas, participants gave fairly consistent responses; in others a greater number of participants reversed their responses. Rehren and Armstrong report that:

depending on the scenario, between 30% and 50% of participants exhibited a rating shift [on a 7 point scale ranging from “Definitely should do it” to “Definitely should not do it”] ($M=49\%$), with the mean magnitude of these shifts being close to two full points (1.93). Between 8% and 20% of participants shifted such that their rating crossed the scale midpoint ($M=14\%$) (2022, p. 10)

Rehren and Sinnott-Armstrong asked study participants who self-reported that they changed their minds to explain why they did so. Of the participants who actually did change their mind on one of the dilemmas (there were some false reports), few reported doing so because they thought about or discussed the dilemmas between the first and second test (2023, p. 15). Rehren and Sinnott-Armstrong conclude that when presented with sacrificial dilemmas in a lab setting, “about 1 in 7 participants would test one way one week, but a different way the next. This level of error does not strike us as acceptable” (2023, p. 16). Here is one explanation that Rehren and Armstrong offer for the inconsistency of their results:

The thought is that for many, perhaps all of our participants, their environment, circumstances or state of mind were different when they participated in the first wave of our study compared to when they participated in the second wave....Another option (though it does not exclude the first option) is that people generally make stable moral judgments, but from time to time, instability (perhaps in the form of distorting influence of situational factors) creeps in for everyone (2022, p. 17)

The empirical research into moral judgment shows us that there is much we don’t understand about which factors influence our judgments. Greene and others have primarily focused on factors intrinsic to the dilemma scenarios, but, as suggested by Rehren and Sinnott-Armstrong, factors about the research participants themselves and the environment in which the study is conducted may be just as important.

In support of this point, Petrinovich et al (1993) found significant gender-based differences in responses to the trolley problem.

On three of the trolley dilemmas women were less likely than men to throw the switch and kill their brother versus five humans (11 % vs. 18%), $x^2(1, N = 387) = 3.9, p < .05$; less likely to do nothing and kill five humans versus a

²⁴ See Sauer (2021) for an argument that reaches a similar conclusion as the one we advance here.

cancer scientist (56% vs. 67%), $\chi^2(1, N = 387) = 4.6, p = .03$; and less likely to push a man in front of a train to save five humans (21% vs. 31%), $\chi^2(1, N = 387) = 4.0, p < .05$ The pattern, then, was for the women to favor a more egalitarian approach—to prefer a lottery versus making a decision on both questions where that was a possible alternative—and less likely to act to kill their brother or to actively push someone in front of a train (p. 473).

Similarly, in Bourget and Chalmers study of the views of philosophers, one of the strongest correlations between gender and philosophical position was on the trolley problem, with female philosophers being significantly less likely to pull the switch in *Switch* (2023, pp. 45-56). Beebe and Sackris (2016) found that female respondents in the general population are less likely to treat moral claims objectively; they also found that when moral judgment study participants know that an issue is controversial in their society, they are less likely to treat the issue objectively. There may well be other factors about respondents that influence their responses. Women and men could be, ostensibly, consciously responding to the same context-based features of the situation, yet nonetheless reaching different verdicts because of factors related to gender. Without investigating factors like these, we cannot be sure what is, or should, influence the judgment of moral dilemmas.

To continue on this theme concerning the make-up of study participants and the context in which the study is conducted, in support of his preference for System 2 verdicts Greene cites a study that shows that when study participants are in a positive or humorous mood they are more likely to give consequentialist responses (2014, p. 703). Presumably, people are not in such moods when they face real moral dilemmas, although college students filling out a survey for money or class credit might be.²⁵ Findings like these should give us a great deal of pause. The context in which we make a moral judgment may have more to do with the verdict we render than any facts about the situation considered in itself. Such evidence should make us skeptical of our ability to find moral truth in the lab; at the very least, it's hard to see how such evidence could be interpreted as confirmatory of any one position.

Let's return to the *Crying Baby* dilemma. Suppose we take Greene's theory at face value and agree with him that it is System 1 processes telling us not to sacrifice our own child, while it is System 2 processes indicating that by doing so we might well save ourselves and many others. If we conducted a study of parents and non-parents, we would probably find some interesting splits in the verdicts.²⁶ Do such factors about the make-up of study respondents matter? Unless we are given good reasons for thinking that they don't, these are the sort of seemingly relevant factors that need to be taken into consideration in a study of moral judgment.

²⁵ Greene cites Valdesto and DeSteno (2006) and Strohminger, Lewis and Meyer (2011) in support of this claim. This evidence also cuts against Greene in another way: According to Kahneman (2011 chapter 5), when we are happy or in a good mood, we are more likely to go with System 1 verdicts. This is additional evidence that utilitarian style verdicts could be generated by either system.

²⁶ We are reminded of Matthew 16:26: "For what is a man profited, if he shall gain the whole world and lose his soul?". For the person who smothered their own child, the number of people they saved, however high, would likely offer them little consolation.

What we have here is a multilayered skeptical problem that seems incredibly difficult to overcome: To have greater confidence in our moral judgments as a result of empirical research, we would antecedently need to know which factors about the participants are relevant/irrelevant and test to see whether subjects are influenced by said factors (e.g., is being a parent relevant? Is being male or female relevant?); the empirical evidence would have to conclusively tell us which system responds to which factors (is considering the perspective of the person tied to the track a System 1 or System 2 consideration?); the empirical evidence would also have to indicate that factors we consider to be morally irrelevant are *not* being taken into consideration (e.g., if gender is morally irrelevant, how do we stop it from influencing judgments?); we would have to already know all of the morally relevant features of a moral dilemma (is the perspective of the person tied to the track morally relevant? Are there other relevant considerations we have yet to even notice?). We have argued that when we take a closer look at the results of a variety of empirical studies we find that we don't have a grasp on any of these facts.

4 In defense of System 1

Thus far we have argued that it is difficult to determine which factors do, or should, influence moral judgment, and that there are likely unknown factors and biases influencing moral decisions whatever cognitive system we take to render a given verdict; we have also questioned our ability to successfully distinguish System 1 from System 2 judgments. However, for all that, we have not called into doubt the more general claim that is central to Greene's normative argument: that we should generally prefer the verdict of System 2 when it comes to certain kinds of judgments. Perhaps, even with the problems we have raised, we still have good reason to prefer System 2 to System 1 for solving certain kinds of moral problems; or perhaps the confounds we have raised can ultimately be ruled out through empirical processes. It is this preference for System 2 judgments and their utilitarian style reasoning that is ultimately central to Greene's argumentative strategy. It is to this claim that we now turn.

The preference for System 2 verdicts when it comes to *Switch*, *Footbridge*, and *Crying Baby* depends on establishing that these really are unfamiliar* problems, and that System 2 really is better for solving these sorts of problems. We can ask two key questions: How do we distinguish between familiar* and unfamiliar* problems; even if we can make this distinction, do we have good reasons for thinking that System 2 outperforms System 1 on unfamiliar* problems, or that System 2 outperforms System 1 more generally when it comes to moral problems?

First, we would need to be able to identify the features that make moral situations familiar*/unfamiliar*. Michael Dale (2020) has already pointed out that this is likely an impossible task. For any given moral decision-context, there are going to be features with which we are both familiar and unfamiliar, and Greene's account of familiarity* is rather vague. As noted above, he says unfamiliar* problems are "problems with which we have inadequate evolutionary, cultural, or personal experience". As we have argued, individuals are very familiar* with comparing sums, which is a

critical part of these dilemmas. Furthermore, humans have a cultural and evolutionary history of making hard choices. In much of the ancient world, infanticide was a common practice carried out for the good of the group, although the practice is now uncommon in much of the developed world (although the moral status of abortion is highly contested). People have lots of cultural experience with trains and trolleys, etc. In short, the distinction just can't be fleshed out in any meaningful way (Dale 2020). This means we can turn to our second key question: Do we have any generalized reasons for preferring the results of System 2, especially when it comes to moral problems?

Greene, and many others who investigate moral judgment using fMRI, do not believe that moral judgment is a distinctive kind or category of judgment. For example, Greene states, "I believe that moral cognition is not a natural kind at the cognitive level" and "So far as we can tell, the field of moral cognition does not study a distinctive set of cognitive processes" (2015, p. 40). Cushman and Young (2011) similarly conclude that moral judgments depend on, or are "derived" from, more general judgment formation processes (p. 1053).²⁷

If moral judgment is no different from "general judgment formation processes" then we would expect it to be just as subject to influence from seemingly irrelevant factors as any other judgment type. For example, Sackris (2021) argues that aesthetic judgment seems to be based on arbitrary factors and that it is impossible to determine whether an art object considered in itself forms the basis of a given aesthetic judgment. Further, Sackris and Larsen (2023) argue that it is unclear how these external factors could ever be overcome, even by art experts.

Just as significantly, if making a moral judgment is just an application of our general judgment forming capacity, then it seems like what we should be studying is our general judgment formation capacity: There really is no use in sticking someone in an fMRI and asking them about moral dilemmas. We might as well ask them about dinner menu items or movie preferences. In other words, if DPT is true, Greene should be asking a much more general question: Which system should we prefer when it comes to making judgments *in general*?

One of the overlooked lessons of Kahneman's classic *Thinking, Fast and Slow* is that although he tends to highlight where our intuitions and heuristics go wrong in our use of System 1, System 1 generally gets things right:

System 1 is indeed the origin of much of what we do wrong, but it is also the origin of most of what we do right—which is most of what we do. Our thoughts and actions are routinely guided by System 1 and generally are on the mark (2011, p. 416).

Notice Kahneman doesn't say "System 1 is generally on the mark for problems experienced in a context with which we are highly familiar". People are constantly facing novel decision-contexts, and if DPT is true they generally navigate

²⁷ See also Young and Dungan (2012) and Decety and Cowell (2014). See Sackris and Larsen (2022) for an extended discussion of the emergence of this position among neuro-scientific researchers into moral judgment.

the world successfully by relying on System 1. If, as Greene believes, moral judgments are merely a sub-species of judgment more generally and not a distinct kind, this gives us good reason to rely on our System 1 judgments when it comes to solving moral problems. That is, let's suppose System 1 really is responsible for deontological judgments. If System 1 generally gets things right and moral judgment is not fundamentally different in kind from general judgments we might form about anything, then Greene should reverse positions: It's the deontological judgments we should trust and the utilitarian judgments we should jettison.

In support of turning Greene's conclusion on its head, we have some reason to think that we ought not rely on our "manual" system in novel circumstances because our "automatic" system often outperforms our "manual" system even in difficult decision-making situations. Consider the following analogy: Czerlinski et al. (1999) had different types of computer analyses compete over prediction accuracy on certain issues while only having access to specific bits of information (the programs competed over which could predict the most accurate dropout rates in Chicago public schools by using only certain bits of information such as test scores, attendance rates, composition of the student population, etc.). Some of the programs were sophisticated and used all the available information (e.g., multiple linear regression analyses), while others were simple and made their decisions based on only one or two bits of information. As most people would expect, the sophisticated analyses did quite well, but, surprisingly, the simple analyses did either just as well or only a little worse (Czerlinski et al., 1999, pp. 102-108). The interesting finding for our purposes is that when these programs used the information that they learned in the first analyses to generalize to new and unknown data sets (i.e., to new and unknown situations), the simple programs consistently performed better than the sophisticated programs (Czerlinski et al., 1999, pp. 108-118).

Martignon and Hoffrage (1999) examined the data, and they concluded that the sophisticated programs were trying to accommodate too much information. Instead of focusing on only the most relevant factors, they made their decision based on all the information available and this led to their predictions being skewed by too many irrelevant factors and statistical noise (i.e., they were overfitting the data). The simple programs, on the other hand, just stuck to one or two factors, and while they sometimes focused on the wrong factors, they were overall more accurate because they usually ended up focusing on fairly relevant factors and weren't being pulled in different directions by irrelevant information.

These studies make an intriguing point about general decision-making capacities. When a system makes a comprehensive analysis of a situation to come to a decision, it's possible for such an analysis to overanalyze and overfit the data, which may lead to a worse decision. Can this understanding be applied to human decision-making? In theory, there is no reason why it couldn't. Like the more "sophisticated" programs, our conscious (System 2) reasoning attempts to consider and weigh many factors, while our intuitive (System 1) processing often reacts to certain, specific stimuli that it has been "programmed" (by culture and/or evolution) to view as important. It seems possible that the more manual cognitive system could suffer from the same shortcoming that the sophisticated computer programs suffered from;

that is, it might focus on too many (irrelevant) factors, and this could lead to a worse decision than relying on simpler (more automatic) mechanisms.

If the simpler mechanisms were consistently focused on the wrong factors, then it is likely that the manual system would be overall more trustworthy, but that is probably not the case with most automatic mechanisms in the human brain; remember, we usually rely on System 1, and we usually make judgments conducive to achieving our goals. Many of the automatic mechanisms (or heuristics) would be the result of long processes of natural selection, and if such a conservative process invests in a mechanism that dictates a certain reaction to specific stimuli, then it is likely that such reactions would often be helpful. Of course, just because a mechanism was helpful in ancestral environments doesn't guarantee that it will be helpful in modern society, but as we are such a variable species that can thrive in all types of climates and conditions, it would make sense for evolution to have equipped us with general mechanisms that assist in making decisions in new and difficult circumstances.

We're not attempting to argue that the automatic system is always, or even generally, more reliable than the manual system; instead that point is that our automatic systems are, considered in themselves, fairly reliable. That is, it isn't abundantly clear when we should prefer the result of one system or another when it comes to solving interpersonal dilemmas, especially if such dilemmas are not different in kind from other problems we face. Consider using an exhaustive check-list to try to decide whether you should marry someone. Although we typically favor this type of analytic approach for difficult decisions, this is typically not the way many people think we should decide upon a romantic partner: In this sort of case, we do think people should 'follow their feelings' or 'go with their gut' even though, for many people, deciding who to marry is a one-off decision they have no experience with.²⁸ There is some evidence that 'going with your feelings' is the right approach for romantic relationships: subjects in a study who simply reported their feelings about their romantic relationships predicted the future of the relationships more accurately than subjects who used analytical reasoning to explain how their relationships were going (Wilson 2002).

Many studies report similar findings (Allman & Woodward 2008; Gigerenzer 2007; Gigerenzer & Gaissmaier 2011; Wilson 2002; Wilson et al. 1993; Woodward & Allman 2007), and they are all consistent with the idea that in some types of difficult decision-making, our automatic system consistently outperforms our manual system. But it's important to understand that this is not just due to the shortcomings of the manual system. Implied in all these studies is the idea that our automatic system is performing better not just because it is focusing on less data, but *because it is picking up on factors that our more manual system is not aware of*. Consider Lewicki's (1986) study, in which subjects were told about the personality traits of

²⁸ Admittedly, in arranged marriages where parents are choosing who their child should marry, they might well make an exhaustive checklist. Whether love-matches are more successful than arranged matches is not a question we will investigate here. It may well be that both have equal levels of success. If so, then we still have no reason to prefer System 1 processes over System 2 processes. Thanks to Rasmus Larsen for raising this point.

people in photographs. What the subjects did not know was that there was a correlation between such traits and the hair length of the people in the photographs. When the subjects were then asked to predict the personality traits of people in new photographs, they correctly extrapolated the correlation between hair length and personality traits, even though they were not consciously aware that they had picked up on the correlation.

In general, this ability of our automatic system to pick up on factors that our manual system is unaware of is a common phenomenon in our species, and it has been consistently corroborated by research (Allman & Woodward 2008; Allman et al., 2005; Klein, 1998; Bechara et al., 1994; Gigerenzer 2007; Gigerenzer & Gaissmaier 2011; Wilson 2002; Woodward & Allman 2007). Of course, System 1 responses can be wrong, but as the research shows, they are often quite accurate in certain (complex) situations, and as mentioned before, there is good reason for this. Because of its slower and more methodical functioning, our System 2 reasoning can only address a limited number of factors in a situation (and some of those, as we have seen, may be irrelevant). Our automatic system, however, has been equipped by both millions of years of evolution and our experiences in life to quickly analyze and weigh potentially important factors that we may not even consciously realize are there.

We could put this point about the liabilities of effortful, System 2 decision-making like this: In effortful decision-making, considering more factors doesn't necessarily lead to better decisions, especially if we are unsure of the moral relevance of the factors that we aim to consciously attend to. And even when individuals attempt to consciously attend to some factors and discount others, this doesn't guarantee that there aren't still other factors that they are yet unaware of that are influencing their judgment.²⁹ If moral judgment is merely a sub-species of our general decision making faculty and not fundamentally different in kind, then we can safely draw the following two conclusions without doing any additional experimental work: (a) moral judgment is just as susceptible to seemingly irrelevant factors as any other judgment type; (b) if System 1 processes are fairly reliable for general decision making, then they are fairly reliable for moral decision making as well.

5 Conclusion

We have argued that moral judgment is subject to a multi-layered skeptical problem using both traditional philosophical argumentation and by pulling together disparate research on moral judgement. We have shown that we don't know what factors moral judgments are in fact responsive to; we don't know what factors they should be responsive to; that if DPT is true we cannot tell which system is yielding which

²⁹ For example, even wine and art experts are highly influenced by contextual factors that they are unaware of when they aim to make conscious judgments of quality. See Cutting (2001); Sackris (2019, 2021); Sackris and Larsen (2023).

verdict; and raised doubts concerning whether we should in fact prefer System 2 verdicts.

We might think that findings like these should make us abandon moral judgment altogether, or at least refrain from making moral judgments whenever possible (Sauer 2021). As we see it, such a position makes little sense if we take the position that moral judgment is no different than any other judgment we might make. As Sackris and Larsen argue in their (2023), that you don't know what factors have influenced your aesthetic judgment doesn't mean that you didn't actually find the art object in front of you to be beautiful. Similarly, if you get a 'bad feeling' as you start to falsify your tax return and as a result judge that you shouldn't go through with your intended act, then you got a 'bad feeling' about your intended behavior that signals to you not to go through with it, whatever feature of the situation might have caused it.

The point is this: If moral judgment is no different from any other judgment, we can't stop making moral judgments any more than we can stop deciding what to wear, or eat, or do for fun. This isn't to say that the empirical investigation of moral judgment is useless or hasn't given us anything. Just the opposite in fact. It has shown us (perhaps unknowingly at the time) that moral judgment is no different from any other value judgment, and just like any other judgment the factors that influence our moral judgment are myriad. Above all, said research has shown us the importance of the context in which we consider a moral problem. But, from our standpoint, empirical research hasn't given us any reason to prefer utilitarian or deontological reasoning processes, or to give up on moral judgment altogether. It has given us reason to be skeptical of our ability to both know what matters morally and know that we are taking the factors that we think matter into account when passing judgment. That is, the empirical research has shown us that we need to take our own moral judgments with a big grain of salt. But we probably should have been doing that anyway: Philosophers from Socrates to Rorty have long been trying to convince us of that very point.

References:

- Allman, J., and J. Woodward. 2008. What are moral intuitions and why should we care about them? *A Neurobiological Perspective. Philosophical Issues* 18 (1): 164–185.
- Bartels, D.M., and D.A. Pizarro. 2011. The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Responses to Moral Dilemmas. *Cognition* 121: 154–161.
- Bauman, C.W., A.P. McGraw, D.M. Bartels, and C. Warren. 2014. Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas. *Social and Personality Psychology Compass* 8 (9): 536–554.
- Beebe, J., and D. Sackris. 2016. Moral Objectivism Across the Lifespan. *Philosophical Psychology* 29 (6): 912–929.
- Berker, S. 2009. The Normative Insignificance of Neuroscience. *Philosophy and Public Affairs* 37: 293–329.
- Bloom, P. 2011. Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review* 99 (2): 26–43.
- Bourget, D., and D.J. Chalmers. 2014. What do philosophers believe? *Philosophical Studies* 170 (3): 465–500.

- Bourget, D., and Chalmers, D. J. (2023). Philosophers on philosophy: The 2020 Philpapers survey. *Philosophers' Imprint*.
- Bruni, T., M. Mameli, and R. Rini. 2013. The science of morality and its normative implications. *Neuroethics* 7 (2): 159–172.
- Carp, J. 2012. The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage* 63 (1): 289–300.
- Cushman, F., and J.D. Greene. 2012. Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience* 7 (3): 269–279.
- Cushman, F., L. Young, and M. Hauser. 2006. The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science* 17 (12): 1082–1089.
- Cutting, J. 2003. Gustave Caillebotte, French impressionism, and mere exposure. *Psychonomic Bulletin and Review* 10: 319–343.
- Dale, M.T. 2020. Neurons and normativity: A critique of Greene's notion of unfamiliarity. *Philosophical Psychology* 33 (8): 1072–1095.
- Dale, M.T., and B. Gawronski. 2023. Brains, trains and ethical claims: Reassessing the normative implications of moral dilemma research. *Philosophical Psychology* 36 (1): 109–133.
- Deming, P., M. Heilicher, and M. Koenigs. 2022. How reliable are amygdala findings in psychopathy? A systematic review of MRI studies. *Neuroscience Biobehavioral Review* 142: 104875.
- Di Nucci, E. 2013. Self-Sacrifice and the Trolley Problem. *Philosophical Psychology* 26 (5): 662–672.
- Evans, Jonathan St. B. T. (2012). Dual-Process Theories of Deductive Reasoning: Facts and Fallacies', in Keith J. Holyoak, and Robert G. Morrison (eds), *The Oxford Handbook of Thinking and Reasoning*, Oxford Library of Psychology (2012; online edn, Oxford Academic)
- Fischer, J.M., and M. Ravizza. 1992. Thomson and the trolley. *Journal of Social Philosophy* 23 (3): 64–87.
- Foot, P. 1967. Abortion and the doctrine of double effect. *Oxford Review* 5: 5–15.
- Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. Viking.
- Gigerenzer, G., and W. Gaissmaier. 2011. Heuristic Decision Making. *Annual Review of Psychology* 62: 451–482.
- Gorr, M. 1990. Thomson and the trolley. *Philosophical Studies* 59 (1): 91–100.
- Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.
- Greene, J.D., F.A. Cushman, L.E. Stewart, K. Lowenberg, L.E. Nystrom, and J.D. Cohen. 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111 (3): 364–371.
- Greene, J.D., L.E. Nystrom, A.D. Engell, J.M. Darley, and J.D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.
- Greene, J. D. (Unpublished). Notes on "The Normative Insignificance of Neuroscience" by Selim Berker. Retrieved from <https://www.joshua-greene.net/misc>.
- Greene, J. D. (2011). Solving the trolley problem. In J. Systema and W. Buckwalter (Eds.), *A Companion to Experimental Philosophy* (1st ed., pp. 175-177). Wiley.
- Greene, J.D. 2014. Beyond point and shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics* 124 (4): 695–726.
- Greene, J.D. 2015. The rise of moral cognition. *Cognition* 135: 39–42.
- Greene, J.D., Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107 (3): 1144–1154.
- Grayot, James D. 2020. Dual process theories in behavioral economics and neuroeconomics: A critical review. *Review of Philosophy and Psychology* 11: 105–136.
- Haidt, Jonathan. 2012. *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Hauser, M., F. Cushman, L. Young, R.K. Jin, and J. Mikhail. 2007. A dissociation between moral judgments and justifications. *Mind and Language* 22 (1): 1–21.
- Kamm, F.M. 2007. *Intricate ethics: rights, responsibilities, and permissible harm*. Oxford University Press.
- Kamm, F.M. 2009. Neuroscience and moral reasoning: A note on recent research. *Philosophy and Public Affairs* 37 (4): 330–345.
- Kamm, Frances Myrna (ed.) (2015). *The Trolley Problem Mysteries*. Oxford University Press.
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

- Klenck, M., and H. Sauer. 2021. Moral judgment and moral progress: The problem of cognitive control. *Philosophical Psychology* 34 (7): 938–961.
- Kohlberg, L. (1984). *The Psychology of Moral Development: The Nature and Validity of Moral Stages. Essays on Moral Development, Volume 2*. Harper & Row.
- Königs, P. 2018. On the normative insignificance of neuroscience and dual-process theory. *Neuroethics* 11 (2): 195–209.
- Königs, P. 2023. Of trolleys and self-driving cars: What machine ethicists can and cannot learn from trolleyology. *Utilitas* 35: 70–87.
- Kumar, V., and R. Campbell. 2012. On the normative significance of experimental moral psychology. *Philosophical Psychology* 25 (3): 311–330.
- Marek, S., B. Tervo-Clemmens, F.J. Calabro, et al. 2022. Reproducible brain-wide association studies require thousands of individuals. *Nature* 603: 654–660.
- McIntyre, A. (2023). Doctrine of Double Effect. In E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* URL = <<https://plato.stanford.edu/archives/sum2023/entries/double-effect/>>.
- Mugg, Joshua. 2016. The dual process turn: How recent defenses of dual process theories of reasoning fail. *Philosophical Psychology* 29 (2): 300–309.
- Mugg, Joshua. 2018. The sound-board account of reasoning: A one system alternative to dual process theory. *Philosophical Psychology* 31 (7): 1046–1073.
- Petrinovich, L., P. O'Neill, and M. Jorgensen. 1993. An empirical study of moral intuitions: Towards an evolutionary ethics. *Journal of Personality and Social Psychology* 64 (3): 467–478.
- Plato. (1981). *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo* (G. M. A. Grube, Trans.). Hackett Publishing Company. (Original work published ca. 385–375 BCE).
- Rehren, P., and W. Sinnott-Armstrong. 2023. How Stable are Moral Judgments? *Review of Philosophy and Psychology* 14 (4): 1377–1403.
- Rorty, R. 1989. *Contingency, irony, and solidarity*. Cambridge University Press.
- Sackris, D. 2019. What Jancis Robinson Didn't Know May Have Helped Her. *Erkenntnis* 84: 805–822.
- Sackris, D. 2021. The Arbitrariness of Aesthetic Judgment. *Journal of Value Inquiry* 55: 625–646.
- Sackris, D. (2023). The disunity of moral judgment: Implications for the study of psychopathy. *Philosophical Psychology*.
- Sackris, D. and Larsen, R. R. (2022). The disunity of moral judgment: Evidence and implications. *Philosophical Psychology*.
- Sackris, D. and Larsen, R. R. (2023). Are there “moral” judgments? *European Journal of Analytic Philosophy*, 19(2).
- Sackris, D., and Larsen, R. R. (2023). The perniciousness of higher-order evidence on aesthetic appreciation. *Dialogue: Canadian Philosophical Review*.
- Sauer, H. 2021. Against moral judgment: The empirical case for moral abolitionism. *Philosophical Explorations* 24 (2): 137–154.
- Schaich Borg, J., C. Hynes, J. Van Horn, S. Grafton, and W. Sinnott-Armstrong. 2006. Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience* 18 (5): 803–817.
- Schwitzgebel, E., and F. Cushman. 2012. Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language* 27 (2): 135–153.
- Schwitzgebel, E., and F. Cushman. 2015. Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition* 141: 127–137.
- Singer, P. 2005. Ethics and intuitions. *The Journal of Ethics* 9: 331–352.
- Small, D.A., and G. Lowenstein. 2003. Helping a victim or helping the victim: Altruism and identifiability. *The Journal of Risk and Uncertainty* 26 (1): 5–16.
- Strohming, N., R.L. Lewis, and D.E. Meyer. 2011. Divergent effects of different positive emotions on moral judgment. *Cognition* 119: 295–300.
- Taurek, J. 1977. Should the numbers count? *Philosophy & Public Affairs* 6 (4): 293–316.
- Thomson, J.J. 1976. Killing, letting die, and the trolley problem. *The Monist* 59 (2): 204–217.
- Thomson, J.J. 2008. Turning the trolley. *Philosophy and Public Affairs* 36: 359–374.
- Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and Biases. *Science* 185 (4157): 1124–1131.
- Valdesolo, P., and D. DeSteno. 2006. Manipulations of emotional context shape moral judgment. *Psychological Science* 17: 476–477.

- Wilson, T. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Harvard University Press.
- Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaran, K. J., & LaFleur, S. J. (1993) Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin* 19: 331-339.
- Woodward, J., and J. Allman. 2007. Moral intuition: Its neural substrates and normative significance. *Journal of Physiology - Paris* 101: 179–202.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.