Moral Constructivism and Empirical Debunking

Jeffrey Brand-Ballard, J.D., Ph.D.
Department of Philosophy
The George Washington University
Washington, DC 20052
(202) 994-6911
jbb@gwu.edu

Abstract

Some philosophers (Joshua Greene, Peter Singer, Walter Sinnott-Armstrong) argue that research on the empirical underpinnings of our moral beliefs supports either skepticism or error theory concerning some of our moral intuitions, such as the widespread intuition that it’s wrong to push the big man off the footbridge in a variation of the trolley problem. I review some of these “debunking arguments” and examine the most prominent rebuttal in the literature: moral constructivism (Neil Levy, Sharon Street, Hallvard Lillehammer, Guy Kahane). Taking T.M. Scanlon as a paradigm constructivist, I agree that constructivism blocks empirical challenges, but I argue that constructivism actually facilitates a different way of empirically debunking some of our moral intuitions, including ones that support deontology. Empirical research can provide clues to what rules would be rejected by reasonable people with evolutionary histories different from ours. Constructivists should recognize that such empirical evidence bears on the moral facts.
Does research on the empirical underpinnings of our moral beliefs call those beliefs into question? Some scientists and philosophers think so, arguing that recent empirical findings support either skepticism or error theory regarding at least some of our moral beliefs. I review some of these arguments and consider the most common rebuttals in the literature, which invoke constructivism. I argue that constructivism actually facilitates a different way of empirically debunking some of our moral beliefs, deontological intuitions being especially vulnerable.

I.

In neuroscientific studies, Joshua Greene and colleagues\(^1\) asked research subjects to contemplate two “trolley problem” scenarios familiar to moral philosophers: the Switch case and the Footbridge case.\(^2\) Subjects contemplating Footbridge exhibited greater engagement of the emotional centers of the brain than did subjects contemplating Switch. Greene explains this finding as follows. During most of human evolution, most violence has been “personal,” so human beings developed a strong emotional aversion to acts of personal violence, such as pushing someone. Flipping a switch in order to turn a trolley, by contrast, is a far less personal means of doing violence. Such impersonal means were unavailable until modern times, so the prospect of flipping the switch doesn’t trigger our evolved emotional aversion.\(^3\)


\(^{2}\)In *Switch*, a runaway trolley is headed for five innocent people who are stuck on the tracks. They will all die if the trolley hits them. There is one innocent person on a side track. If you flip a switch, the trolley will turn onto the side track, killing one but sparing five. Most people believe that it is permissible to flip the switch. In *Footbridge*, you and a man are standing on a footbridge over the tracks. His body is large enough to stop the trolley if you push him onto the tracks. He will die but the five others will be spared. Many people believe that it would be morally wrong for you to push the big man onto the tracks. The trolley problem originates with Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967): 5-15.

\(^{3}\)Greene has since acknowledged that his original characterization of the personal/impersonal distinction was inadequate. Joshua D. Greene, "Dual-process Morality and the Personal/Impersonal Distinction: A Reply to
Peter Singer, along with Greene, uses this neuroscientific/evolutionary story to argue that our moral intuition against pushing the big man off the footbridge\(^4\) isn’t reliable enough to use as a starting point for reflective equilibrium.\(^5\) Singer and Greene offer a *debunking explanation*: a causal explanation of why we have the intuition that doesn’t entail its truth. Evolution, they claim, favored individuals to whom pushing the big man *seems* wrong, whether or not it actually *is* wrong.

Singer is no moral skeptic or error theorist. He doesn’t insist that all moral intuitions are unreliable or false. On the contrary, he relies on intuitions such as his belief “that five deaths are worse than one.”\(^6\) Greene, too, supports consequentialism on intuitive grounds.\(^7\) Both maintain that they rely on intuitions that are more “rational,” “reasoned,” or “cognitive” than the intuition about pushing the big man, and that the former are therefore more reliable than the latter.\(^8\)

II.

Folke Tersman responds with debunking explanations of Singer’s own intuitions, such as the intuition that an individual has a reason to promote her own welfare. Evolutionary psychology can explain why Singer has that intuition, without assuming its truth. Debunking

---

\(^4\)Unless otherwise noted, “pushing the big man” denotes the act of pushing him off the footbridge to his death in order to stop the trolley and spare the five others in the Footbridge case described.


arguments lead not to consequentialism, Tersman suggests, but to value skepticism. 9  Guy Kahane advances a similar objection to Singer. 10

Hallvard Lillehammer offers a different rebuttal to debunking arguments. 11 These arguments, he claims, presuppose that moral facts are response-independent. A causal account of a moral belief doesn’t undermine its truth if its truth is a function of, for example, the dispositions of informed agents. Several other philosophers make similar points. Neil Levy, Sharon Street, and Mark Timmons each insist that empirical debunking only works if moral facts are taken to be mind-independent. 12 Antirealism insulates our moral beliefs from empirical challenges. It’s emerged as the rebuttal of choice for “success theorists.”

III.

It’s puzzling that philosophers who’ve advocated skepticism or error theory on empirical grounds have left themselves vulnerable to constructivist rebuttals. Constructivism is hardly an obscure or newfangled metaethical position. It’s been in the mainstream for decades. 13 Empirically-based arguments should confront it directly. I’ll argue that empirical results could, in principle, have implications for constructivists. However, I agree that empirical results provide no support for global skepticism or error theory until constructivism is refuted (if that’s

even possible). At least, the results to date provide no such reason, and no future results are likely to provide one.¹⁴

Levy and Timmons cite T.M. Scanlon as a paradigm constructivist,¹⁵ so I’ll take his contractualism as representative. Scanlon believes that an act is wrong if and only if someone could reasonably reject any set of rules that permits it. That’s what makes an act wrong, Scanlon believes.¹⁶ According to contractualism, pushing the big man onto the trolley tracks in Footbridge is wrong if and only if someone could reasonably reject any set of rules that permits it. A set of rules that permits pushing includes a pushing rule.¹⁷ Many reasonable people today would, indeed, reject pushing rules, so contractualists could conclude that pushing the big man is wrong.¹⁸ Many of these same people would accept a set of rules that permits flipping the Switch.¹⁹ An evolutionary psychologist could offer a causal hypothesis explaining these divergent beliefs. Different types of violence have played different roles in human evolution. Greene’s explanatory hypothesis uses the personal/impersonal distinction, but I won’t rely on it because it can’t explain some newer published data.²⁰ Instead of explaining the data in terms of distinctions that lack obvious moral significance, as Greene does, let’s appeal directly to the distinctions that deontologists themselves have drawn, such as the distinction between

---

¹⁴Again, assuming that constructivism is a live option.
¹⁶Scanlon, What We Owe to Each Other, p. 189.
¹⁷More precisely: a “lesser evils” exception to the general rule forbidding interpersonal violence.
instrumental harm – harming someone as a means to an end – and all other harm, which is non-instrumental.\(^\text{21}\) An example of instrumental harm: one caveman kills another in order to consume his body, share his flesh with his hungry kin, and/or make tools from his bones. An example of non-instrumental harm: a caveman digs a trench in order to deflect an impending flood from his cave. The only effective path for the trench leads to his neighbor’s cave, which will now flood. He has nothing against his neighbor, but someone’s cave is going to flood and he’d rather it not be his.\(^\text{22}\)

Evolutionary psychologists could suggest that, throughout human evolution, there have been more opportunities for individuals to benefit by inflicting instrumental harm than non-instrumental harm. In some cases, of course, instrumentally harming promotes inclusive fitness. If the caveman kills someone for food, but saves his two children who’d otherwise starve, then that’s a net gain for his genes. But if he kills someone just to get his children an extra snack, and one of them is killed in retaliation, then that’s a net loss.\(^\text{23}\) A community of individuals who made a practice of harming one another whenever they saw a short-term benefit wouldn’t thrive in the long term. So groups with an aversion to inflicting instrumental harm enjoy a selective advantage over those without such an aversion.

By contrast, opportunities to enhance fitness by inflicting non-instrumental harm have been relatively rare. Such opportunities mainly arise when one faces a threat that can be deflected and one has the practical wisdom and ability to deflect the threat, thereby endangering someone else as an effect or aspect of the deflection.\(^\text{24}\) That kind of flood just doesn’t strike very often. Therefore, groups with an aversion to inflicting non-instrumental harm wouldn’t enjoy


\(^{\text{22}}\)Assume, if you wish, that a flooded cave spells death.

\(^{\text{23}}\)Of course, we also need an evolutionary account of retaliation behavior, but I’ll leave that to others.

\(^{\text{24}}\)“Effect or aspect” comes from Kamm, *Intricate Ethics*, p. 164.
much of an advantage, if any, over those without such an aversion. This aversion to instrumental harm persists into the present. Many reasonable people, therefore, reject pushing rules. No similar aversion to non-instrumental harm has evolved. Reasonable people today understand that flipping the switch saves lives, so they accept rules permitting it. Even the hapless man onto whom the trolley is turned can’t reasonably reject the flipping rule – or so the contractualist story goes.

We need some real evolutionary psychology, here, not armchair speculation, but I think such a story could be made plausible. Would it have philosophical implications if we had compelling empirical evidence for it? Could accepting it weaken one’s reason to reject the pushing rule? Here’s an argument for the affirmative. Instrumental harm occasionally promotes inclusive fitness, but too often it doesn’t. Primitive humans couldn’t identify with sufficient accuracy the fitness-enhancing occasions for instrumental harming. Therefore, evolution selected for a categorical aversion to doing it. Our emotions tell us there’s something intrinsically wrong about doing it, but really there isn’t. As more advanced beings, however, we can sometimes identify reliably the unusual cases, such as the Footbridge, in which instrumental harm is socially beneficial. So we should disregard our evolved aversion to instrumental harm when making moral judgments about those special cases, much as we should disregard our evolved desire for fatty foods, now that gorging on fat doesn’t enhance fitness or long-term welfare.

Contractualists will object that an evolutionary story, however compelling, is merely a causal explanation of why people reject the pushing rule. It doesn’t undermine their practical

---


reasons for rejecting it. Reasonable people reject the pushing rule because it’s practically reasonable to do so. In support of this claim, contractualists could disparage the pushing rule in various ways. They’ll say that the rule, or the pushing permitted thereby, “fails to respect rational nature,” “expresses disrespect for humanity,” “compromises our inviolability,” “undermines human dignity,” “permits harming someone as a means to an end,” or “permits treating human beings as less than ends-in-themselves.” Natural selection may explain how we came to recognize these facts, but it doesn’t falsify the associated judgments. By analogy: there’s surely an evolutionary story that explains why our brains can recognize arithmetic facts, but that story can’t falsify our arithmetic judgments. Even if scientists prove that reasonable people reject the pushing rule “because” categorical condemnation of instrumental harm was fitness-enhancing, that doesn’t imply that it’s ever actually permissible to inflict instrumental harm even when, as in Footbridge, doing so happens to minimize deaths.

IV.

Let’s grant the contractualist that the pushing rule disrespects rational nature, or compromises our inviolability, or however she wants to put it. Let \( P \) denote the proposition that the failure of the pushing rule to respect rational nature constitutes a reason to reject it. If the contractualist asserts \( P \), then my next question for her is: what sort of fact is \( P \)? One could give either a realist or an antirealist answer. What I’ll call a \textit{disrespect-realistic} would say: “\( P \) is a

\[\text{For such language see any number of deontological authors. Kamm, Intricate Ethics; Judith Jarvis Thomson, "The Trolley Problem" in Rights, Restitution, and Risk (Cambridge, MA: Harvard University Press, 1986); Elizabeth Anderson, Value in Ethics and Economics (Cambridge: Harvard University Press, 1993).}\]


Suppose two people both want to abide by whatever contract they would negotiate in reasonable circumstances – but they fail to agree on which circumstances would be reasonable. Perhaps . . . they are in genuine disagreement about what circumstances would be reasonable. If so, old metaethical questions return in new guise. What is at issue in disputes over what is reasonable? What would justify accepting one answer and rejecting another?
mind-independent moral fact of the universe. It’s not a function of anyone’s actual or hypothetical mental states.”

Disrespect-realism entails that pushing the big man has intrinsic properties that make it reasonable to reject the pushing rule. Whereas, the disrespect-antirealist says: “To say that P is true is just to say that a reasonable person would consider pushing the big man to be disrespectful to such a degree, or in such a way, that she would reject the pushing rule.”

Empirical findings could, in principle, make trouble for either of these answers. Such studies reveal more than how we actually are; they can help us to infer how we might have been under different circumstances, which is also an empirical matter. Imagine a hypothetical society – Paxia – in which 95% of the inhabitants have substantially internalized rules forbidding instrumental harm. Perhaps Paxians simply lack the mental or physical ability to deliberately harm one another. Maybe they don’t understand each other’s physical vulnerabilities well enough to exploit them. They don’t ordinarily harm others instrumentally. They sanction rule-breakers. However, the Paxians also accept the pushing rule. Fortunately, their nervous systems are hard-wired with a specialized and highly reliable “emergency detection” capacity. When someone (anyone) in Paxia is standing on a footbridge and sees a runaway trolley headed for helpless people, she suddenly understands that they’re in danger and that pushing the big man off the bridge will spare their lives. This practical wisdom only comes to her in actual emergencies, when pushing the big man really would save lives. If she’s a good samaritan, then she acts on

---

29 Don’t confuse a disrespect-realist with a realist about the property of all-things-considered wrongness. Contractualists can’t be realists about the latter, but they can be disrespect-realists.

30 This is just an example of what a disrespect-antirealist might say.

31 Compare H.L.A. Hart:
There are species of animals whose physical structure (including exoskeletons or a carapace) renders them virtually immune from attack by members of their species and animals who have no organs enabling them to attack. If men were to lose their vulnerability to each other there would vanish one obvious reason for the most characteristic provision of law and morals: Thou shalt not kill.
this knowledge. But the wisdom is temporary and localized: it doesn’t carry over to the rest of her life. She never mistakenly concludes that there’s danger when there isn’t. Nor does she become more prone to violence, generally. She never gets the idea to push one of her enemies off a cliff, for example. All Paxians have this psychology. These facts are common knowledge in Paxia. Paxians know that they’re like this, know that other Paxians know, and so on. They evolved this way over millennia, isolated from the rest of humanity. Although Paxians inflict instrumental harm more often than we do, they don’t otherwise harm one another more often than we do, and fewer Paxians than Americans are fatally struck by trolleys in Footbridge cases.

However, if pushing the big man is intrinsically disrespectful, as per disrespect-realism, then pushing the big man is wrong, even in Paxia. Therefore, the disrespect-realist is committed to holding that reasonable Paxians would reject the pushing rule, just as we do, despite their different evolutionary history. In support of this claim she could argue as follows. Suppose we describe Paxia to reasonable people in our society and ask them: if you lived in Paxia, would you reject the pushing rule? Reasonable people in our society would answer affirmatively, predicts the disrespect-realist. Therefore, pushing the big man is wrong, even in Paxia.

I’m actually not sure how real people would answer this question. Someone should do a study. If most reasonable Americans answered negatively, then the disrespect-realist would be embarrassed. But even if all Americans answered affirmatively, I wouldn’t infer that any reasonable Paxian would do so. At least, I wouldn’t infer this if I had an evolutionary account of our psychology and a contrasting account of Paxian evolution. The Paxians evolved under conditions different from ours, so I wouldn’t assume that their answers would resemble ours.

The contractualist might reply that she’s read my description of Paxia and its history. She can imagine how she’d feel, think, and act if she were a reasonable Paxian with a Paxian

---

history, and she’s still sure she’d reject the pushing rule. But reading a description of Paxia may not be enough, especially if it’s my thin, cartoonish portrayal. Reading a memoir of a soldier’s life isn’t like serving in combat, oneself. Third-person knowledge is often inferior to first-person knowledge. Reading my description of Paxia imparts only third-person knowledge of Paxian circumstances. The Paxians would have first-person knowledge.33

My argument thus far entails that, according to contractualism, it’s wrong for Paxians in Paxia to push the big man if and only if reasonable Paxians could reject a rule that permits it.34 Of course, we can’t visit Paxia and ask. At best we can extrapolate, estimate, speculate, and predict. Here’s where empirical studies of the real world become relevant to contractualism. Such studies could inform our “predictions” of what reasonable Paxians would think if they existed. Empirical studies can’t, however, challenge contractualism itself. Antirealism may have other problems, but vulnerability to empirical refutation isn’t one of them. Nevertheless, contractualism buys its immunity to empirical refutation at a price. Contractualism holds that the hypothetical responses of certain human beings constitute the moral facts. That’s why empirical explanations of our moral beliefs don’t undermine them. But contractualism doesn’t hold that the moral facts depend on what we happen to believe.35 It holds that the moral facts depend on what well-informed, reasonable people would believe. That’s not always going to be us. The more scientists learn about the causal origins of our moral beliefs, the more accurately they’ll be able to extrapolate to what other human beings, who are otherwise like us, would believe under various counterfactual conditions. The moral facts under those conditions will be a function of what rules reasonable humans under those conditions would reject.

34The antecedent of this biconditional isn’t a report of conventional Paxian morality. It’s a normative claim, like the claim that the ancient Aztecs were wrong to practice human sacrifice, despite the fact that they believed otherwise.
35Cf. normative cultural relativism.
V.

Why should we care what Paxians would think if they existed? One reason is purely theoretical: if reasonable Paxians wouldn’t reject the pushing rule, then pushing the big man isn’t, after all, wrong in all empirically possible societies. It isn’t, as many deontologists claim, intrinsically wrong.

I predict that reasonable Paxians wouldn’t reject the pushing rule. If I’m wrong, then deontology scores big. Unlike us, the Paxians weren’t subjected to selection pressures that would lead them to support the no-pushing rule. If, nevertheless, even reasonable Paxians would reject the pushing rule, then the judgment that pushing the big man is wrong starts to look as much like an “objective moral fact” as anything contractualism countenances. My point is that the Paxians’ own answer matters, and it’s an empirical issue – a complex one – what that answer would be.

There’s also a more practical reason to care what Paxians would think. If reasonable Paxians wouldn’t reject the pushing rule, then it’s not irrational for us to wish that we were more like them in the relevant respects. We might have good reason to wish to be more like them: better at identifying real emergencies and predicting consequences accurately, less tempted to harm one another for selfish or foolish reasons. If we discovered that it were possible for us to become more like the Paxians in those respects, then we might wish to do so. We’d want to determine how similar to the Paxians we’d have to become in order for it to make sense for us to abandon our no-pushing rule.

To many, the very idea will seem perverse: wanting to become enough like the Paxians someday to abandon our no-pushing rule. But I think constructivists can’t vindicate such opposition. They could, however, argue that becoming sufficiently similar to Paxians is
psychologically impossible for us. If so, then trying is futile, and even consequentialists should simply resign themselves to the no-pushing rule as the best rule our psychology allows. But I honestly don’t know how similar to Paxians we could become. Neither do constructivists. I’d like to know. It’s an empirical question.

VI.

Some researchers oversell the philosophical implications of empirical results, suggesting that they support global error theory or skepticism. Success theorists block that challenge by endorsing constructivism. But this move enables empirically-minded philosophers to mount a different attack, challenging the central deontological claim that pushing the big man is intrinsically wrong. With the right empirical results on their side, they could, in principle, defend a localized error theory, specifically targeting deontological restrictions. When moralists embrace constructivism, ironically, they make empirical results more relevant than ever to that potentially revisionist program.