MIDWEST STUDIES IN PHILOSOPHY

Midwest Studies in Philosophy, XLIII (2019)

Debunking, Vindication, and Moral Luck*

DANIEL STATMAN

1. INTRODUCTION

The last decade or two have witnessed a comeback of debunking arguments in morality.[1] In particular there is an increasing interest in evolutionary debunking arguments[2] that attempt to undermine moral realism by arguing that insofar as the evolvement of morality was determined by its contribution to survival, there is no reason to think that it tracks moral *truth*.[3] What will occupy me here, however, are debunking arguments which lead to a more positive conclusion, namely, arguments that rely on a debunking strategy in order to offer a breakthrough in some longstanding moral debate. As far as I know, such a strategy has been used with regard to two moral debates: (i) the debate between utilitarians and deontologists; (ii) the debate between supporters of moral luck and its opponents. Attempts have been made to debunk deontology, thereby supporting utilitarianism, and to debunk moral luck, thereby supporting the no-luck position. In both cases, the assumption is that "if there is a tension among a set of beliefs and we find out that one subset is unjustified, then this lends support to the other subset."[4]

*I'm grateful to David Heyd and Saul Smilansky for helpful comments on earlier drafts. Thanks also to Andy Khoury for his generous editorial assistance.

1. See Nichols, Timmons, and Lopez (2014).

2. See Kahane (2011): "Whether in implicit or explicit form, evolutionary debunking arguments are increasingly being put to use in recent normative ethics."

3. The worry about the effect of evolutionary theory on morality is part of a more general worry about the reliability of our epistemic faculties given the truth of evolutionary theory plus the assumption that there is no God. See Plantinga (1993).

4. Kumar and May (2019).

DOI: 10.1111/misp.12113
© 2019 Wiley Periodicals, Inc.

My focus here will be on (ii), namely, on attempts to make progress in the moral luck debate by relying on descriptive assumptions taken from psychology and from evolutionary theory. Most of these attempts utilize such assumptions against moral luck,[5] though some do so to defend it.[6] The former believe that empirical studies *debunk* moral luck; the latter believe that they *vindicate* it. I argue that the strategy fails in both its forms. Since the fundamental difficulties in this strategy also apply to the debate between utilitarians and deontologists, I start with a brief discussion of the debunking strategy in this debate and then turn to the use of debunking—and vindicating—arguments in the context of moral luck. I had my first go at this strategy fifteen years ago when it was making its first steps.[7] Here I would like to strengthen my case against it and to discuss other attempts that have appeared since then to use empirical research in order to solve the moral luck problem.

If I'm right, then progress in psychology or in evolutionary theory probably cannot solve normative problems such as whether luck affects blameworthiness or whether deontological constraints have intrinsic moral force. To solve such problems, we have no choice but to go into the philosophical debate.

In the past, I used to think that although debunking arguments are of no help in solving normative problems, they can at least serve as a basis for theories of error.[8] The thought was that while pointing to the assumed unreliability of the epistemic process that led somebody to believe p is rarely sufficient to discredit p, if joined by independent arguments against p, it, nevertheless, can explain how intelligent and well-intended people could err and believe p. I now doubt whether this move can work either.

2. THE FAILURE OF DEBUNKING DEONTOLOGY

In a series of studies, Joshua Greene relies on two kinds of evidence in order to debunk deontological intuitions and pave the way for the acceptance of utilitarianism.[9] One is taken from neuroscience. In a pioneer fMRI study, Greene showed that when subjects think about the standard trolley case (diverting the trolley to a side track in order to save five, knowing that this will kill one worker on the tracks) thereby typically making a utilitarian judgment, the parts of the brain associated with cognition and rational processing are activated. By contrast, when they consider the fat man case (killing one person intentionally in order to save five), the parts of the brain associated with emotions are activated, but less so with the more rational parts. This seems to show that deontological intuitions are grounded in emotion while utilitarianism is grounded in reason. And this supposedly indicates that we should listen to the voice of reason—to utilitarianism—instead of allowing ourselves to follow our emotions and accept deontological constraints.

5. See Domsky (2004), Levy (2016), Khoury (2018), Young, Nichols, and Saxe (2010).
6. Kumar (2018); Nichols et al. (2014).
7. See Statman (2005), which responded to Domsky, "There is No Door" (2004).
8. Statman (2005), 436.
9. See Greene (2013), mainly chs. 4–5 and 9.

The other kind of evidence comes from evolutionary theory. Our instinctive, deontological responses must have come from somewhere, and if we accept the theory of evolution, then they must have developed as a result of evolution. This means that our immediate negative reaction to the pushing of the fat man onto the tracks must be a result of a long evolutionary process. Greene identifies the process as follows: For our ancestors, who lived in small societies, it was crucial as far as possible to refrain from attacking other people, especially people in their own communities. Such attacks carried great risk for the attackers because retaliation against them would have been very harsh. Since, in such societies, such attacks almost always meant attacks in close proximity, that is, pushing somebody, hitting him, stabbing him and so on, our ancestors developed a kind of an internal alarm system that warned them of the possible danger every time they considered carrying out an attack on others. The internal alarm made perfect sense for our ancestors: it protected them from the potentially lethal results of engaging in violence, and it contributed to a more peaceful and cooperative society with better odds of survival. However, argues Greene, circumstances have changed, and this internal alarm is no longer relevant in a world in which people can harm each other in all kinds of ways, only few of which involve actual personal contact such as pushing or stabbing. The goal of reducing harm and improving the odds of survival is, of course, still crucial, but the way to achieve it is by utilitarian reasoning that ignores constraints that no longer make sense.

Both arguments assume that the epistemic process by which we came to believe in deontological constraints is unreliable, either because it is rooted in emotion, or because it was created to cope with a danger that the alarm no longer tracks very well. Hence the belief in such constraints is debunked and the belief in utilitarianism confirmed.

I turn now to some critical comments which will be helpful when I turn later to the moral luck debate. First, the explanation for the evolvement of our intuition against pushing the fat man is highly speculative. Needless to say, there is no independent evidence that could support it, particularly evidence showing that societies whose members were armed with this alleged internal alarm managed to survive while societies that failed to develop it gradually disappeared. We also have no independent knowledge about the moral intuitions of our ancestors. Greene is simply assuming that if we, today, have an aversion to pushing the fat man in order to save five lives, our ancestors must have had it as well.

This leads to a general worry about speculations regarding the evolution of our moral intuitions. We don't start by studying pre-historic societies and then infer from the evidence that they have evolved by natural selection, but rather simply *assume* that the main (or maybe the only) mechanism underlying the evolvement of the social, emotional and rational features of human beings is natural selection, and then conclude that all traits held by human beings today must have evolved in order to raise their odds of survival. The best way to see how speculative such explanations are is to think of

evolutionary explanations that would have probably been presented if the assumptions about the current makeup of human beings turned out to be mistaken. Assume, for instance, that a future study contradicts Greene's assumption about an innate aversion to the use of upfront violence. This, of course, would not be taken as a *refutation* of evolutionary theory, but as suggesting that we are in need of an alternative explanation to that proposed by Greene. Here is one that might be proposed in this imaginary state of affairs: In a world of scarce resources, it was crucial for animals, and later for human beings to develop a propensity to violent aggression. It enabled them to obtain food and water by force whenever possible, and it provided them with a way of defending themselves and of deterring potential attackers from taking their goods. Both goals, especially the defensive one, required a determinate readiness to use upfront physical force when necessary, because hesitance in the use of force would have been far less effective in achieving deterrence. Just as an immediate violent response to threats improves the odds of survival for animals (that probably don't have the internal alarm system ascribed to humans), it seems to have improved the odds of survival among humans as well. According to Azar Gat, "aggression has always been a *major* option, and thus very close to the surface and easily triggered,"[10] and in his view "strong selection pressures over many millions of years have made it so."[11] The speculative nature of the evolutionary explanations for moral intuitions will play an important role in my argumentative strategy below.

Second, since support for utilitarianism is based on intuition just like support for deontology, it is equally susceptible to debunking arguments. Here's a possible one: The principle of utility made perfect sense within small communities struggling for survival. The utility referred to, however, was that of a specific society, not of the entire world. When the human population grew, the principle was re-interpreted to apply to the overall utility of all human beings. Arguably, while the principle of utility made sense—and was implementable—for our ancestors, this is no longer the case for us today. Why—and how—should we care about the *overall utility*?

If you think that this evolutionary speculation is too wild, I could make up others. Once again, the point is that making up such stories is too easy and doesn't really advance the normative debate between utilitarians and deontologists. If you are convinced that utilitarianism is true, you could always come up with some story to explain the evolvement of the allegedly mistaken deontological intuition—and the other way round. Clearly, then, these stories do not decide the debate; they enter the scene after the debate has already been decided—to the satisfaction of one side or of the other—on other grounds.

All this leads to the suspicion that when people point to defective epistemic processes that allegedly lead their opponents astray, *they themselves are biased in doing so*. This suspicion gets support from a body of psychological research showing that people tend to be oblivious to biases in themselves

10. Gat (2006).

11. Ibid.

while readily recognizing the impact of biases on others.[12] This fact, together with the ease of making up evolutionary stories, strengthens the main point of this section to the effect that such stories can play a very limited role in deciding normative debates like the one between utilitarians and deontologists. This paves the way for my discussion of the use of empirical research in the context of the moral luck debate, to which I now turn.

3. DEBUNKING MORAL LUCK; GENERAL DISCUSSION

Before I turn to discuss some recent attempts to use the debunking strategy in order to resolve the moral luck debate, let me offer a few general observations. First, this strategy is only used in order to refute (or, more accurately: to undermine) the claim that moral luck exists,[13] not in order to refute the opposite claim expressed by 'the control condition' to the effect that one is never blameworthy for what is beyond one's control.[14] I mention this fact since it's all too easy to construct evolutionary debunking arguments and, *a priori*, there is no reason to expect that they should be used only by one party to the debate and not by the other. While the control condition is often described as a "compelling abstract principle" challenged by "intuitions about particular cases,"[15] this description is biased in favor of the control condition, granting it a more privileged epistemic status. In fact, to say that some principle is 'compelling' is no different than to say that we have a strong intuition in its favor[16]—precisely the kind of intuition that we have in those examples proposed by proponents of moral luck in order to *undermine* the control condition. Ultimately, then, the problem of moral luck arises because of a seeming conflict between strong intuitions on both sides: an intuition supporting the control condition and intuitions about particular cases that support the existence of moral luck.[17] The point is that if the origin of some moral judgments (or ultimately all of them[18]) is based on direct intuition and not on some rational process, we are forced to ask where this intuition came from and—if we accept something like Darwinism—this is the same as asking what evolutionary process led to it.

12. See Pronin, Gilovich, and Ross (2004).

13. See Domsky (2004), Levy (2016), Khoury (2018).

14. See also Levy (2016: 128), who says, referring to the question of whether the two negligent drivers should be seen as equally blameworthy or not that "neither option seems fully satisfactory because each leaves a powerful intuition unaccounted for." For the claim that both sides to the moral luck debate can play the debunking-explanation game, see Enoch and Guttel (2010).

15. Kumar (2018), 1.

16. When asked whether blameworthiness should ever hinge on the luck of how things turn out, 90% of respondents said that it should not. See Lench et al. (2015).

17. As David Enoch notes, the intuitions on both sides of the debate seem "close to moral bedrock," which leads to a "nagging worry" about whether there is some deeper intuition or premise to rely on in order to decide the debate. See Enoch (2015).

18. See Khoury (2018: 1369 n. 38), who notes that "it is not clear there is any rational alternative to appealing, at bottom, to intuition in moral theory."

To put the point in other words, when the idea of evolution is used to explain the human brain, behavior, psychology, culture, and so on, the only evolutionary mechanism relied on is natural selection and not other mechanisms like mutation or genetic drift. The presumption is that all traits present in humans today have evolved because, at some phase of human evolution, they raised the odds of survival among our ancestors.[19] But if that's the case, it must apply to the control condition as well. We must assume that belief in the control condition gave our ancestors some kind of evolutionary benefit, thanks to which it raised their odds of survival. But that immediately makes this belief susceptible to the standard evolutionary debunking argument according to which there is no reason to assume a correlation between what is beneficial and what is morally true.

What story could be told to explain the evolvement of (belief in) the control condition? Here's one that comes to mind: In a society in which people were held blameworthy (by themselves and by others) for bringing about bad results regardless of whether they were responsible for them or not, people developed a kind of a fatalistic approach, saying that "it's all a matter of luck anyhow." As a result, they did not make a serious effort not to be agents of harm. Because this approach turned out to be harmful, it gave way and was replaced by a more beneficial one, according to which responsibility is limited to what is under our control. If you are not convinced by this story, a few minutes of creative thought can produce an alternative one. The ease of producing debunking arguments for both sides of the debate raises serious doubt as to how helpful these arguments might be in deciding the debate.

Note, secondly, that the debunking efforts are aimed at undermining only one type of moral luck, namely, luck in the consequences of one's actions ("resultant luck"),[20] and even within that category usually only in cases in which the bad consequences are due to *negligence*, not in those due to deliberate decisions that turned out to be mistaken (such as signing a peace treaty that led to a disaster) or in those due to good or bad luck in the execution of a deliberate attempt (the successful assassin versus the unsuccessful one).[21] This double restriction of (the objects of) the debunking strategy is ad hoc[22]

19. For the sake of the present discussion we can bracket the question of whether traits evolved *because* they raised the odds of survival or whether they evolved randomly and then survived due to their evolutionary advantage.

20. See Domsky (2004), Levy (2016), Khoury (2018). In the estimation of Robert Hartman, "Moral Luck and the Unfairness of Morality," *Philosophical Studies* (forthcoming), "the most popular position in the moral luck debate is that circumstantial and constitutive moral luck do exist but that resultant moral luck does not." He argues that this *Asymmetric View* is untenable.

21. See Khoury (2018: 1371 n. 46), who rejects Domsky's debunking argument because it is limited to cases involving outcomes brought about unintentionally. Indeed, I see no basis for Nichols et al.'s claim (2014: 163), that cases of moral luck "typically" involve negligence (see also Khoury's reference to Domsky 2004: p. 445).

22. See Enoch and Guttel (2010: 376–77).

and undermines its ambition to solve *the* problem of moral luck. Even if the strategy works in cases of negligence, one can't assume *a priori* that it works in other types of moral luck. For instance, there is no basis for assuming that the evolutionary process that led to the ascription of more blame to unlucky agents than to lucky ones is the same leading to the ascription of more blame to agents who took bribes than to those who would have taken bribes if offered, but luckily were not.

All this leads to the worry that there is something unreliable in the debunking project itself, namely, that the ascription of biases, especially those rooted in evolutionary explanations, to the opposing view (the view that endorses moral luck) is itself biased. One worries that (i) unreliable epistemic processes are searched for (or made up) only in the case of the pro-luck view and not in the case of the anti-luck one, and (ii) regarding the former a big fuss is made of (assumed) biases in one sub-kind of moral luck in order to jump to the conclusion that luck never affects morality (more accurately, that no judgment expressing moral luck is reliable).[23]

These reflections confirm a point made earlier to the effect that in order to establish that some epistemic process is defective (and hence the judgment based on it unreliable), we must first make substantive assumptions that do not rely on such assumed defects. In the context of the moral luck debate, this means that we first need to make up our minds about the relevant ethical and meta-ethical questions, and only then can we turn to speculations about the factors that assumingly led our opponents astray. More generally, in order to determine that some epistemic process is defective, we must have an independent way of distinguishing between true and false judgments in the relevant domain, on the basis of which we can then judge that some processes are reliable (because they tend to lead to true judgments) while others are defective (because they do not).[24]

A possible response would be that some epistemic processes, like tossing a coin, are so evidently defective, that even without an independent way of distinguishing true and false judgments in some domain, one can know that judgments originating from such processes are unreliable. Hence, contrary to what was just suggested, one *can* establish the unreliability of judgments assuming moral luck—on the basis of observations about their dubious source—without first confronting the philosophical arguments for and against the influence of luck on morality.

The problem with this response is that the identification of such evidently defective epistemic processes that assumingly led somebody to believe p is rarely possible. It is true that "if I decide on the basis of a coin toss to believe that the number of stars in Andromeda is even, then my belief is unreliably formed,"[25] but, first, I'm skeptical as to whether one could inten-

23. See especially Domsky (2004) who presumed to have "finally solved the problem of moral luck."

24. See Rini (2016).

25. Hill (2016), n. 11.

tionally decide to believe p,[26] especially on the basis of facts which evidently do not support p. Second, even if forming beliefs on the basis of evidently irrelevant facts like (the result of) coin tossing were possible, it would be very hard to point to actual processes of this kind. At any rate, none of the epistemic processes that are deemed unreliable by proposers of debunking arguments in morality are of this kind: (i) The holders of the debunked positions don't *intentionally* rely on a defective source, (ii) The claim that their holding of this position was unconsciously caused by such a source is rather speculative; and (iii) the defect itself is often not evident. Therefore, if, to debunk moral positions directly without relying on substantive arguments, one needs to point to defective processes like deciding on the basis of coin tossing, debunking arguments would very rarely get off the ground.

Maybe the test for epistemic defectiveness can be weakened to include processes which are not as manifestly defective as tossing a coin. For instance, maybe the very fact that some position is not a result of reasoning is sufficient to undermine it, with no need to go into the philosophical arguments for or against moral luck. The problem with this move is that it leads to the debunking of almost all of our moral views. Jonathan Haidt famously said that "intuitions come first, strategic reasoning second,"[27] and if intuitions do not come from reasoning, they can only come from non-rational processes—evolution, peer pressure, parental influence, and so on—which, on this proposal, would necessarily make them all unreliable. But that would render the detailed attempts to debunk moral luck redundant. The philosophers involved in such attempts try to show that *specific* intuitions, those assuming moral luck, are unreliable.

In the next section I discuss a recent attempt by Neil Levy to debunk moral luck, which illustrates some of the difficulties I pointed to in the current section. In the following sections, I discuss attempts to use empirical studies in order to confirm—or to vindicate—moral luck.

4. LEVY ON HOW EVOLUTION DEBUNKS MORAL LUCK

Neil Levy's 2016 paper is a good illustration of an attempt to solve (or dissolve[28]) the problem of moral luck by debunking the pro-luck position and showing that "the intuitions generated by focusing on the consequences of actions are unreliable" (137). Let me introduce his argument and then explain why I think it fails.

Like others, Levy focuses on resultant luck, more specifically on cases of negligence. Relying on Fiery Cushman's research,[29] he observes that blame judg-

26. See Alston (1989). But see Steup (2017).

27. Haidt (2012).

28. In Levy's rather technical terms, to *solve* the problem of moral luck is to explain why we experience a conflict in intuitions in its regard. To *dissolve* it is to provide good reasons for preferring one set of intuitions to the other, which is done by debunking (belief in) the latter.

29. See mainly Cushman (2008).

ments are independently sensitive to mental states (beliefs, desires, intentions) and to the consequences of actions. Because we are sensitive to mental states, we tend to think that negligent drivers whose relevant mental states are similar are morally on par regardless of the contingent results of their driving (the one hitting a child, the other not). And because we are also sensitive to consequences, we tend to judge the driver who hit a child more harshly than the one who luckily did not.

This observation, however, falls short of explaining the inner conflict we sense when we face such cases. If mental states *and* consequences matter for blameworthiness, why not simply give each its due weight? Why the sense of *paradox*?

To answer these questions, Levy offers an evolutionary explanation for this dual nature of blame judgments. Blame, he suggests, "is a response that protects against free riding and the collapse of cooperation" (131). Such protection was crucial for the survival of our ancestors and it was better achieved by blaming people for their bad intentions and desires than for the contingent consequences of their actions. What makes people defect from cooperative arrangements are their problematic mental states and capacities; hence, it is these states and capacities that society has come to target when ascribing blameworthiness.

How, then, did sensitivity to *consequences* evolve? In line with other epistemic reductionists,[30] Levy suggests that because of the high correlation between causing harm and intending harm, our ancestors developed an automatic tendency to assign importance to consequences when ascribing blame. Ultimately, then, our sensitivity to mental states and our sensitivity to consequences grew out of the same aim, namely, to improve cooperation and minimize free-riding. Blame for mental states achieves this aim more directly, blame for consequences more indirectly, as a proxy for the former.

But if we care about consequences only because that's usually a good way to learn about agent intentions, why do we care about them even when the picture regarding these intentions is evident? Why are we so torn in the face of dilemmas where it is stipulated that two agents differ morally only in the results of their actions, not in their mental states? Levy's answer relies on a general view according to which psychological processes tend to be "encapsulated," that is, "insensitive to personal-level information" (134). Once our ancestors developed the concern for consequences in blame judgments—albeit as a mere epistemic tool—it came to retain an intuitive appeal even in circumstances in which it was irrelevant. In other words, it became an automatic response, efficient in the many cases in which taking consequences into account was appropriate, while misfiring in the few ones in which it was not.

However, to use Joshua Greene's metaphor, notwithstanding the advantage of taking pictures in an automatic mode,[31] it makes no sense to do so in circumstances where the manual mode is expected to yield much better pictures. Similarly, notwithstanding the benefit of automatically treating bad consequences as a proxy to bad intentions, when it is obvious that the intention of one agent

30. See, e.g., Richards (1993), ch. 9.

31. Greene (2013), 133.

is no worse than that of the other, it makes no sense to impose upon them both different levels of blame even if, due to bad luck, one led to worse results. The pull we find within ourselves to do so should not be trusted.

To conclude, according to Levy, the evolutionary origin of blame judgments lies in the need to protect against free-riding. To do so most effectively, such judgments evolved to track those mental states that encouraged defecting from social cooperation. Because of the high correlation between causing harm and intending harm—and the difficulty of tracing harmful intention when no harm was produced—the bringing about of bad consequences came to be considered an independent source for imposing blame. But, of course, it is not. Evolutionary theory thus helps us to understand both how the puzzle of moral luck arises and how it can be dissolved.

I turn now to some critical comments. First, as Levy himself concedes (2016: 134), the story he tells about the adaptive function of blame is rather speculative. There is, of course, no evidence to indicate that a (very) long time ago human beings lacked the emotion of blame or the practice of blaming, which some of them then developed in order to improve the level of their social cooperation—and that indeed it did, which is why we, today, feel blame and impose blame on others. Levy simply assumes that any emotion or practice we find today (i) must have evolved over time and (ii) must have evolved because of some advantage it yielded for survival. But even if these two premises are granted, the emotions and practices of our ancestors might have been advantageous in ways that we can't even imagine today, hence all explanations for their emergence are highly speculative.

Moreover, one can never know for sure whether a human trait reached the twenty-first century *because of* its adaptive advantage or in spite of its *lack* of such advantage. Think of a basketball team that won a game. It obviously won thanks to its players, to how they scored, took rebounds, defended their net and so on. But that doesn't mean that the game was won thanks to the positive contribution of *each* of the players. One player might have had a bad day and, as a result, might have made a negative contribution to her team's game. So the team won (or "survived"), but it did so in spite of that player, not thanks to her. Similarly in the evolution of human traits; some must have evolved because they raised the odds of survival, but not necessarily all of them. Moreover, even if all human traits originally evolved because of their evolutionary benefit, as circumstances changed—and they changed dramatically throughout natural and human history—some traits might have lost their benefit to humankind and may have even become a burden.[32]

32. This difficulty is exacerbated if one accepts Talbott's (2015) that argues "norms that appear pointless or even disadvantageous in isolation may be part of a system of norms that is greatly advantageous to the group," and furthermore that even otherwise arbitrary social norms are usually beneficial "because they help to distinguish between those to whom cooperation is owed (the in-group) and those to whom it is not (the outsiders)." On this view, how could we know whether some norm was disadvantageous "in isolation," but nonetheless contributed to the group system "as a whole," or whether it was advantageous in itself, thus providing a seemingly stronger reason to think that it should continue to be adhered to today as well?

Once again, to appreciate the speculative nature of Levy's evolutionary story here, note how easy it is to construct alternative ones. Consider the following: Maybe the independent importance assigned to consequences in the practice of blame evolved as a result of the (assumingly beneficial) evolvement of the belief in mythology. This belief included the idea of the magical, namely, the idea that doing certain things—sacrificing an animal, uttering a prayer or a curse, desecrating a holy place, and so on—bring about good or bad consequences regardless of the agent's intentions or the quality of his will. Maybe this is the evolutionary source of the intuition that the very doing of certain things, such as killing a human being, is morally bothersome or contaminating independently of the extent to which it manifests bad mental states. Once again, if you don't like this story, I could come up with another one. They all seem highly speculative and it's really hard to decide between them.

Nonetheless, for the sake of argument let's accept Levy's story which assumingly entails the reliability of blame judgments that respond to (negative) mental states and the unreliability of those that respond to (bad) consequences. The problem is that the same reason that undermines the reliability of the latter also undermines the reliability of the former. In Levy's view, the intuition about the independent contribution of bad consequences to blameworthiness is unreliable because of the contingent and imperfect correlation between harmful consequences and harmful intentions. But the same argument applies to mental states as well. According to Levy, (negative) mental states evolved as the criterion for blameworthiness because of the need to protect against free-riding. Yet the causal relation between the practice of blame and the reduction of free-riding is also merely contingent. We could think of cases in which self- or other-blame makes no contribution to social cooperation, for instance when one torments oneself with blame on one's deathbed. Our strong and stable intuition to impose blame on people every time their mental states are morally deficient is, therefore, unreliable.

But, again, I might be wrong. So let's grant the adaptive function of blame judgments directed at mental states and also the epistemic benefit of regarding consequences as an independent factor in determining blameworthiness. The natural question is why this should matter for our *current* practice of blame judgments. Prima facie, it would seem to matter if we are consequentialists, but actually it does not. The fact that some human trait or social arrangement played a role at some stage of human evolution does not entail that it is still beneficial today, in circumstances so different and so removed from those in which this trait or this arrangement evolved. One could think of many attitudes and practices that were common to most of humanity for thousands of years, hence must have had some adaptive value at some phase of human history—but, nonetheless, are widely regarded today as harmful. Consider the following example. Males have the inclination to utter strong swear words more than females have. According to a recent study, this inclination has an evolutionary explanation. Briefly, the idea is that—

> From an evolutionary perspective humans are most likely to survive if they have access to resources; if they can defend their resources and protect their families; and if they can attract or gain access to mates. Aggressive male behavior seems to have evolved to support the human race in achieving all of these primary goals.[33]

Yet arguably today such aggressive male behavior expressed in the use of strong swear words or of other milder forms of violence is more harmful than beneficial.

Back then to the practice of blame judgments. To establish the current benefit of this practice today, it is of very limited help to speculate on ancient pre-history. What is needed instead is evidence about this benefit *today*.[34] Methodologically, providing such evidence would be a serious challenge, but, until the challenge is met, we are in the dark about the actual contribution of blame judgments (that assume or that deny moral luck) to social cooperation.

Finally, and most importantly, for *non*-consequentialists, even if some practice were both beneficial for our ancestors and still beneficial for society today, this fact still falls short of establishing that we ought to maintain it. To utilize the above example of male aggression, even if such aggression were overall beneficial to society, many would insist that it ought not to be followed. Put simply, if, for non-consequentialists, consequences in the *present* world are insufficient to determine the morality of practices, all the more so with the assumed consequences for our *ancestors* of maintaining certain practices.

In the conclusion of his paper, Levy says that "evolution has disposed us, for good reasons, to be implicitly committed to a theory of blame according to which blame ought to be commensurate to actors' mental states," and to regard bad consequences as an independent factor in assigning blame. But even if he is right in assuming that these commitments have good *evolutionary* reasons—reasons to explain why we humans evolved as we did—he has provided us with no *moral* reasons to subscribe to these commitments. There is nothing in his argument to rule out the possibility that the moral truth regarding blameworthiness diverges from the guidance of the intuitions given to us by evolution. In particular, it is possible that the actual consequences of actions make a genuinely independent contribution to blameworthiness, as argued by supporters of moral luck. Levy has, therefore, failed to unravel the moral luck puzzle; failed to give us good reasons for preferring the pro-luck intuition over the anti-luck one.[35]

33. Güvendir (2015).

34. Recall Greene's point mentioned earlier according to which while the special aversion to causing upfront harm, like pushing or stabbing, made evolutionary sense (= was beneficial) in the world of our ancestors, it no longer makes sense for us today. For us, it makes no sense to think that pushing a person to his death is any worse than pressing a button that will make him fall and die.

35. Levy (2016), 37.

5. KHOURY; PSYCHOLOGICAL DEBUNKING OF MORAL LUCK

As a matter of psychological self-report, Andrew Khoury finds it "intuitively obvious that resultant moral luck does not exist."[36] However, he realizes that others have the opposite intuition, hence his very pointing to a strong anti-luck intuition is not sufficient to decide the matter. So he sets himself "to defend the rejection of moral luck,"[37] and the way he does so is by debunking the intuition that moral luck exists.

This time, the proposed debunking is not evolutionary but psychological. The famous hindsight bias is assumingly responsible for the fact that we care about outcomes and not only about the quality of the agent's will. This bias misleads us to think that if some state of affairs materialized, it must have been—or at least could have been—foreseen by the agent. Hence, the fact that we regard harmful agents as more blameworthy than agents who were fortunate not to cause harm is not due to the intrinsic contribution we assign to outcomes in determining blameworthiness, but to the mistaken thought that the quality of the will which ended up causing harm was worse than that of the will which did not. We unconsciously take the bad results as evidence for the defective will of the agent for which she *is* assumingly responsible. Thus, "the best explanation for the existence of the intuition that [moral] luck exists is one that accords it no evidentiary weight."[38] The idea that belief in moral luck is debunked by the hindsight bias has been utilized by other writers as well,[39] but the differences between them does not concern us here.[40]

Does Khoury provide supporters of moral luck with a reason to withdraw their support? I don't think so. First, the defense of the anti-luck position can't rely on the intuition that moral luck does not exist because, as mentioned above, supporters of moral luck have their own—opposite—intuition. The problem is that without this reliance it is hard to see how the purported defense might work. To see why, consider an analogous move by those *supporting* moral luck, aimed at debunking the anti-luck position. What's the best explanation for the holding of this (assumingly) counter-intuitive position? Maybe the fact, mentioned by Bernard Williams in his classic paper, that this position—

> has an ultimate form of justice at its heart, and that is its allure. Kantianism is only superficially repulsive—despite appearances, it offers an inducement, solace to a sense of the world's unfairness.[41]

According to this debunking argument, the reason that people accept the counter-intuitive view that no moral luck exists is that they are desperate for

36. Khoury (2018), 1369.
37. Ibid.
38. Ibid.
39. See Royzman and Kumar (2004); Domsky (2004).
40. See Khoury 1371: n. 46.
41. Bernard Williams, "Moral Luck," in Statman (1993), 36.

a solace to the world's unfairness. In other words, what drives the (assumingly) mistaken rejection of moral luck is plain wishful thinking.[42] We *want* the world to be fair; judgments based on luck appear unfair; hence we bring ourselves to believe that luck cannot play a role in moral judgments.[43]

This wouldn't be a very convincing defense of the moral luck position. For the same reason, the opposite debunking, the one proposed by Khoury, doesn't work either. For Khoury, the best explanation of the intuition that there is no moral luck is that it tracks the truth about the role of luck in morality, whereas the best explanation of the opposite belief is the hindsight bias. For proponents of moral luck it is the other way round: the best explanation of the intuition that moral luck exists is that it tracks the truth about the role of luck in morality, whereas the best explanation of the opposite belief is wishful thinking and the just-world bias. The lesson, once again, is that we first need to settle the philosophical debate between the conflicting intuitions, and only afterwards can we come up with speculations to explain why the mistaken party failed to see the light.

Second, the hindsight bias is insufficient to explain how outcomes influence blame judgments even when the quality of the will is clear, for instance, when we stipulate that two drivers are equally negligent. This has to do with the fact that the hindsight bias affects us unconsciously, not when we are presented with unambiguous data about the probability of some event. But, with moral luck, even when convinced that two agents are equal in terms of their inner world, many people—and many justices as well[44]—believe that outcomes make a difference. Furthermore, we often recognize this fact in advance even before we know the actual outcome. A politician signing a peace treaty often realizes at that very moment that if the peace is successful he'll be a hero and remembered as such, but if it fails, he'll be damned forever both by others and by herself. Her realization that her moral status depends on outcomes over which she has no control has, therefore, nothing to do with the hindsight bias.

Third, we mentioned above Cushman's study according to which blame judgments manifest sensitivity to (bad) outcomes independently of what the outcomes might teach us about the agent's mental state, which seems to refute Khoury's proposal that outcomes influence our blame judgments only because they (wrongly) change our view about the quality of the agent's *will*. Khoury handles this objection by proposing that Cushman's results about the appropriateness of blame are concerned with the question of whether *overt*

42. This wishful thinking might well have to do with the "just world fallacy" (see, e.g., Lerner 1980).

43. This alternative debunking argument—against the anti-luck position—might take care of Khoury's objection to Enoch and Guttel (Khoury 2018: 1372 n.48) to the effect that they have failed to lift the burden of articulating an alternative argument to the one proposed by critics of moral luck.

44. Elsewhere I showed that although, according to Israeli penal law, criminal (failed) attempts deserve the same punishment as successful ones, in actual cases, courts impose harsher penalties in cases where the attempts are successful. See Statman (2010), 11–34 (Heb.).

or *expressed* blame is overall justified or not—not the question of whether the agent was *blameworthy* or not. That is why Cushman's findings do not undermine his anti-luck position. There is nothing inconsistent in arguing that while blameworthiness is insensitive to outcomes, expressed blame (and punishment) are.

I can't deny that this is a possible interpretation of Cushman's findings, but it does seem to me somewhat forced, given the fact that the subjects were explicitly asked "How much blame does [agent] *deserve*?"[45]; to deserve blame and to be blameworthy are the same thing.

Fourth, as argued by Enoch and Guttel, the hypothesis about the effect of the hindsight bias on moral luck intuitions is ultimately an empirical one, and to date there is no empirical data in its support.[46]

To conclude, then, I don't think that the debunking argument that Khoury proposes provides the defense he promises for his rejection of resultant moral luck. Ultimately, what underlies this rejection is his sense that it is "intuitively obvious" that such luck doesn't exist, together with the general observation that in moral theory there is probably no alternative to an appeal to intuition.[47]

6. KUMAR ON HOW EVOLUTIONARY THEORY VINDICATES MORAL LUCK

Victor Kumar also believes that the key to making progress in the problem of moral luck lies in recent empirical work, in psychology and in evolutionary theory. A helpful way to introduce his view is to look at the point at which he parts company from Levy. Both agree that we should try to locate the evolutionary origin of blame judgments. They also agree that this origin must have the form of natural selection, namely, that the practice of blame judgments as we know it evolved because of the benefits that accrued to our ancestors. Finally, they agree about the kind of social benefit that was advanced by the practice of blame, namely, the reduction of anti-social behavior and the encouragement of cooperation. They disagree, however, about the form of blame that could have brought about such evolutionary benefits. In Levy's view, the best way to reduce anti-social behavior is to impose blame on people with bad intentions because it is these intentions that underlie their anti-social behavior. That's why blame judgments focus primarily on (bad) mental states and only secondarily, as an epistemic heuristic, on the actual consequences of these states in terms of harm to other people. As we saw, this led Levy to the denial of moral luck.

By contrast, in Kumar's view, the best way to reduce anti-social behavior is to focus on outcomes, which are much easier to identify than intentions. Blaming for intentions would mean holding many people mistakenly

45. Cushman (2008), 358, italics added.

46. Enoch and Guttel (2010: 382).

47. Khoury (2018), 1369 n. 38.

blameworthy and failing to hold others blameworthy although they deserve it. Thus, "assigning responsibility partly on the basis of outcomes is more reliable—more reliable than assigning responsibility purely on the basis of opaque intentions—and thus is able to regulate behavior more effectively" (Kumar, 12). A blame-imposition system that is vulnerable to luck is, therefore, more beneficial to society, and hence more justified.

The problem with this defense of moral luck is that it sounds like a plain consequentialist justification, and the role of the evolutionary story becomes unclear. What Kumar proposes as a response is the following: every time it is pointed out that consequentialism leads to the justification of some seemingly problematic rule or practice, consequentialists immediately find a sophisticated way of showing that the rule or practice is actually ruled out by consequentialism itself. The impression is that they can justify any practice or rule on the basis of its direct or indirect contribution to better overall results, which raises the suspicion that "whatever our practices had been like, we would have been able to find some consequentialist justification for them … their ubiquitous availability casts doubt on any particular application" (2018: 15). The role of the evolutionary argument in Kumar's theory is to dispel this suspicion. It seeks to show that a consequentialist justification of the existing practice of blame is not a post-hoc rationalization of a practice that evolved for different reasons, but of a practice that grew out of, or was *based on* (ibid.) consequentialist considerations.

I don't find this answer convincing. First, as argued above, evolutionary explanations (at least with regard to human emotion, behavior, social structures, etc.) suffer from a similar "ubiquitous availability" and, in that sense, are no better than consequentialist ones. Second, as there is such minimal independent evidence about the actual emotions and behavior of our ancestors and of the transitions that occurred over time in them, to claim that what drove the evolvement of emotion and behavior (including the practice of blame) was their adaptive function—their overall benefit—is just to project backward the explanation one adheres to *today* about the function of, for instance, the practice of blame. Because one assumes that in the contemporary world this practice has good social consequences, one concludes that this must have underlain its initial evolvement thousands of years ago. So, contrary to Kumar, evolutionary explanations don't provide consequentialism with independent evidence that could confirm the hypothesis concerning the social benefits of the practice of blame, but the other way round; *because* this practice is already perceived as based on its positive social consequences, evolutionary explanations of its evolvement are structured in the same way. The suspicion, therefore, has not been dispelled.

It's worth noting in this context the ambivalence of Kumar's proposal that our practices of blame and punishment are *based on* the consequentialist considerations that justify them (2018: 15, italics added). While "based on" often conveys a relation of justification, like in "this conclusion is based on three premises," this can't be the case here because what grounds conclusions

are propositions and not natural events like those assumed by evolutionary explanations. So Kumar is probably using "based on" in a descriptive sense, as in "the despotic regime is based on the people's fear," or "the movie is based on a true story." But if that's the case, then, once again, the fact that the practice of blame was selected because of its evolutionary advantage thousands of years ago says very little about its advantage today and, at any rate, what is *advantageous* is not necessarily *morally true*. Kumar himself concedes that "some human dispositions and practices may be the products of evolution by natural selection but are decidedly unjustified, e.g., ethnic biases against outgroup members" (2018: 18). But this is precisely what opponents of moral luck think of the practice of blame imposition which is sensitive to luck, namely, that it is decidedly unjustified.[48] Just as ethnic bias can't be justified by the fact that it probably evolved by natural selection, moral luck can't be justified this way either.

Kumar coins a new term—"vindication"—to describe cases in which "one and the same account of moral change combines explanation with justification."[49] The general idea is that "moral progress tends to occur when people carry out successful 'experiments in living'… Thus, when a moral change is progressive, the justification for it is usually reflected in its etiology."[50] In Kumar's view, moral luck is of this kind,[51] namely, its etiology—its evolutionary origin—vindicates its justification. This novel idea of vindication merits more discussion that can be offered here, so I'll limit myself to three brief comments: (i) It's a bit unclear in what sense moral luck can be seen as a case of "moral *progress*" to which the vindication model could be applied, or even as a case of moral *change*. As argued above, we have no evidence for the premise that societies first had a "Kantian" practice of blame, focusing, as it were, only on intentions, which was later replaced by a practice that assumed moral luck. (ii) Since, by their nature, evolutionary explanations are "consequentialist," in the sense of grounding the evolvement of emotions and behavior in their advantage, that would mean that only consequentialist justifications could be vindicated (when using evolutionary arguments), which raises the worry that the argument is just a sophisticated rationalization of consequentialism. (iii) The idea of vindication might work in those cases in which we have reliable knowledge about the learning process that led to some belief and we have access to the considerations that went into that process. But, as emphasized above, nothing of this sort applies to the explanation of how human beings came to believe that outcomes play a role in determining blameworthiness. Even worse, the thought that (unconscious)

48. See, e.g., Khoury (2018: 1369), who contends that it is "intuitively obvious that resultant moral luck does not exist." See also Enoch's estimation (2015: 48), that deniers of moral luck "will insist on the control condition and find it almost unbelievable that some philosophers—even some good philosophers—deny it."

49. Kumar (2017).

50. Ibid., 128.

51. Ibid, section 5.

consequentialist considerations underlay the evolvement of this belief among our ancestors is a mere projection to the past of a view concerning the function of blame judgments in the present. As it *assumes* the latter, it can't vindicate it.

7. NICHOLS ET AL. ON THE ENTRENCHMENT OF MORAL LUCK JUDGMENTS

Like Kumar, Shaun Nichols et al. also think that psychological research supports the idea of moral luck rather than debunks it. They start with a relatively weak condition for what makes the holding of normative commitments "legitimate," by which they mean something like reliable, or epistemically responsible. Against the idea that such legitimacy requires that the commitment be grounded in a rational and reliable process, they argue for what they call "ethical conservatism" which says that "certain normative commitments need not flow from rationality to retain their authority" (160). Which commitments have this status? Those that are "entrenched in our psychology," namely, (i) are not the products of inferences from other norms of facts and (ii) are rooted in human emotion (161). According to Nichols et al., moral luck judgments, that is, judgments that ascribe blame (also) on the basis of outcomes, satisfy both conditions, hence are "legitimate," even though they are the product of patently a-rational and a-reliable processes.

Let me first express some doubt about the idea of ethical conservatism. If a commitment is not inferred from factual beliefs, where could it come from? The answer proposed by Nichols et al. is that it is "rooted in human emotion" (2014: 161). But while Hume would have probably stopped there, noting that "our examination of causes must stop somewhere,"[52] after Darwin this is no longer an option. Instead, we must ask what factors underlay the evolvement of this emotion and the standard reply is that the evolvement was driven by natural selection. But why assume that natural selection is a reliable process for acquiring truth in the domain of blame and punishment (or, for this matter, in any other domain)? Contrary to Nichols et al., the less inferential a commitment is and the more rooted in emotion, the *less* reliable it seems.

That as it may be, the proposal is that psychological research, mainly by Cushman and by Young et al.,[53] shows that bad outcomes (of negligent agents) lead directly to harsher blame, hence the additional blame imposed on the driver who hit a pedestrian in comparison to the one who didn't is not an inference from other facts. In particular, it is not an inference from the hitting driver's flawed desires or intentions. Condition (i) is thus satisfied. Moral luck judgments satisfy condition (ii)—the rootedness in emotion—as well, hence the relevant normative commitments are entrenched in the sense alluded to above.

52. David Hume, *An Enquiry into the Sources of Morals*, ch. 2. n. 8.

53. Cushman (2008) and Young et al. (2010).

The problem with this explanation, say Nichols et al., is that sometimes emotions exert a distorting bias on our judgments. Therefore, in order to establish the legitimacy of the commitment to moral luck, this worry needs to be removed. Here is how they suggest we do it: In the authors' view, "one strategy of exposing bias is to see whether people withdraw their judgment under full information" (2014: 173). If they do, this is evidence that the judgment was premature; that it was a result of some kind of bias. An empirical study conducted by the authors confirmed that even under full information subjects assigned critical weight in their blame judgments to the unfortunate outcomes of negligent behaviour, namely, they stuck to their moral luck judgments. When, however, the unfortunate outcomes were a result of non-negligence—of good reasons—subjects assigned much less weight to the outcomes in their blame judgments. This shows that judgments of outcome-based blame are sensitive to whether or not the agent was negligent, which means that the moral luck responses cannot be dismissed "as the product of a distorting general emotional bias" (2014: 174).

I will grant that there might not be an *emotional* bias here, but, because the distinction between (the attitude to) unfortunate outcomes resulting from negligence and those resulting from 'good reasons' is irrational, *some* bias is quite clearly at play. The irrationality has to do with the inconsistency in the weight assigned to unlucky outcomes in the case of good reasons in comparison to that assigned to them in the case of bad ones. For deniers of moral luck, fairness requires that people be held responsible only for what is within their control, and since the outcomes of their actions are not, they cannot be held responsible for them. That applies to good results as it applies to bad ones; to good mental states (reasons, intentions) as to bad ones.[54] By contrast, for supporters of moral luck, outcomes do count morally (namely, correctly affect blame judgments) whether the mental states that lead to them (or to the actions that produced them) were good or whether they were not. Note that one of Thomas Nagel's examples of resultant (bad) moral luck is of a politician who conscientiously and responsibly makes a momentous decision like going to war, or signing a peace treaty, which then leads to a disaster.[55] Hence, if blame judgments are much more affected by outcomes in the case of bad reasons (negligence) than in the case of good reasons, this seriously undermines their reliability. They might still be "entrenched in our psychology," but this psychology now seems untrustworthy in the domain of blame judgments.

8. CONCLUSION

The fact that the intuitions we have in favor of or against some moral principle, or some concrete case, are a result of an evolutionary process governed

54. For the symmetry between moral blameworthiness and moral praiseworthiness in the sense relevant to the moral luck debate, see for instance Khoury (2018), 1358.

55. Thomas Nagel, "Moral Luck," in Statman (1993), 61–62.

by natural selection says very little about their truth or their reliability. From an evolutionary perspective, *all* intuitions must have evolved this way, which is compatible with the thought that many of them, like the intuition that serious criminals ought to be put to death (common in almost all societies up to the twentieth century[56]) should be rejected either because they are no longer beneficial or because they are morally wrong even *if* overall beneficial. Evolutionary explanations, then, have a very limited role to play in determining moral questions.

This applies to the problem of moral luck as well. Both the pro-luck and the anti-luck positions are based on strong intuitions about concrete cases and about the control condition. The fact that such intuitions must have evolved as a result of some evolutionary advantage is insufficient to help us decide which is more reliable than the other. Empirical studies in psychology and in evolution can, it appears, neither debunk nor vindicate moral luck.[57]

REFERENCES

Alston, William. 1989. *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, NY: Cornell University Press.

Cushman, Fiery. 2008. "Crime and Punishment: Differential Reliance on Causal and Intentional Information for Different Classes of Moral Judgment." *Cognition* 108: 353–80.

Domsky, Darren. 2004. "There Is No Door: Finally Solving the Problem of Moral Luck." *Journal of Philosophy* 101: 445–64.

Enoch, David. 2015. "Moral Luck and the Law." *Philosophical Compass* 5: 42–54.

Enoch, David, and Ehud Guttel. 2010. "Cognitive Biases and Moral Luck." *Journal of Moral Philosophy* 7: 384–85.

Gat, Azar. 2006. *War in Human Civilization*. Oxford: Oxford University Press.

Greene, Joshua. 2013. *Moral Tribes*. London: Atlantic Books.

Güvendir, Emre. 2015. "Why Are Males Inclined to Use Strong Swear Words More Than Females? An Evolutionary Explanation Based on Male Intergroup Aggressiveness." *Language Sciences* 50: 133–39.

Haidt, Jonathan. 2012. *The Righteous Mind*. London: Penguin Books.

Hill, Scott. 2016. "From Isolation to Skepticism." *Erkenntnis* 81(3): 649–68.

Kahane, Guy. 2011. "Evolutionary Debunking Arguments." *Noûs* 45: 103–25.

Khoury, Andrew. 2018. "The Objects of Moral Responsibility." *Philosophical Studies* 175: 1357–81.

Kumar, Victor. 2017. "Moral Vindications." *Cognition* 167: 124–34.

———. 2018. "Empirical Vindication of Moral Luck." *Noûs*. Advance online publication. https://doi.org/10.1111/nous.12250

Kumar, Victor, and Joshua May. 2019. "How to Debunk Moral Beliefs." In *Methodology and Moral Philosophy*, ed. Jussi Suikkanen and Antti Kauppinen. London: Routledge.

Lench, Heather, Darren Domsky, Rachel Smallman, and Hathleen Darbor. 2015. "Beliefs in Moral Luck: When and Why Blame Hinges on Luck." *British Journal of Psychology* 106: 273–87.

Lerner, Melvin. 1980. *The Belief in a Just World: A Fundamental Delusion*. New York: Plenum.

56. See https://www.mun.uzh.ch/en/session/Past-Sessions/Capital-Punishment.html.

57. I am, thus, less optimistic than Enoch and Guttel (2010, 386) about the prospects that the study of cognitive biases (together with philosophical argumentation) "may yet advance the moral-luck debate."

Levy, Neil. 2016. “Dissolving the Puzzle of Resultant Moral Luck.” *Review of Philosophy and Psychology* 7: 127–39.

Nichols, Shaun, Mark Timmons, and Theresa Lopez. 2014. “Using Experiments in Ethics: Ethical Conservatism and the Psychology of Moral Luck.” In *Empirically Informed Ethics: Morality Between Facts and Norms*, ed. Markus Christen, Carel van Schaik, Johannes Fischer, Markus Huppenbauer, and Carmen Tanner, 159–78. Berlin, Germany: Springer.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.

Pronin, Emily, Thomas Gilovich, and Lee Ross. 2004. “Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others.” *Psychological Review* 11: 781–99.

Richards, Norvin. 1993. “Luck and Desert.” In *Moral Luck*, ed. Daniel Statman. Albany: State University of New York Press.

Rini, Regina A. 2016. “Debunking Debunking: A Regress Challenge for Psychological Threats to moral Judgment.” *Philosophical Studies* 173: 675–97.

Royzman, E., and R. Kumar. 2004. “Is Consequential Luck Morally Inconsequential? Empirical Psychology and the Reassessment of Moral Luck.” *Ratio* 17: 329–44.

Statman, Daniel (ed.). 1993. *Moral Luck*. Albany: State University of New York Press.

———. 2005. “Doors, Keys and Moral Luck: A Reply to Domsky.” *Journal of Philosophy* 102: 422–36.

———. 2010. “The Criminal Attempt After Amendment 39 to the Penal Law in Light of the Problem of Moral Luck.” In *Trends in Criminal Law: Ten Years After Amendment 39 to the Criminal Law*, ed. Eli Lederman, Keren Shapira-Etinger and Shai Lavee. Ramat Gan: Bar-Ilan University Press.

Steup, Matthias. 2017. “Believing Intentionally.” *Synthese* 194: 2673–94.

Talbott, William. 2015. “How Could a ‘Blind’ Evolutionary Process Have Made Human Moral Beliefs Sensitive to Strongly Universal, Objective Moral Standards?” *Biology and Philosophy* 30: 700–01.

Young, Liane, Shaun Nichols, and Rebecca Saxe. 2010. “Investigating the Neural and Cognitive Basis of Moral Luck: It's Not What You Do but What You Know.” *Review of Philosophy and Psychology* 1: 333–49.