Noûs

NOÛS 53:4 (2019) 987–1007
doi: 10.1111/nous.12250

Empirical Vindication of Moral Luck

Victor Kumar
Boston University

Abstract

In resultant moral luck, blame and punishment seem intuitively to depend on downstream effects of a person's action that are beyond his or her control. Some skeptics argue that we should override our intuitions about moral luck and reform our practices. Other skeptics attempt to explain away apparent cases of moral luck as epistemic artifacts. I argue, to the contrary, that moral luck is real—that people are genuinely responsible for some things beyond their control. A partially consequentialist theory of responsibility justifies moral luck. But this justification is no mere rationalization of the status quo. Recent experimental and evolutionary work on punishment and learning suggests that the very same reasons that justify moral luck have also shaped the evolution of our luck-sensitive moral practices.

Imagine that you have been entrusted with the care of a friend's child. You're preparing to drive the child home, but you're distracted and neglect to secure her safely in a car seat. On the way, you're struck by another driver and the child is gravely injured. How should you feel? Intense guilt and remorse would be natural. How should the child's parents react? It would, of course, be reasonable for them to be angry with you and to blame you severely.

Now imagine that I have also been entrusted with the care of a friend's child, and that I too neglect to secure her safely in her car seat. Fortunately, however, I happen not to be hit by another driver. It's only after I pull into the driveway that I realize my mistake. How should I feel in this situation? Self-reproach is appropriate, but it would be an overreaction to feel the same intense guilt and remorse you feel in your situation. Were the child's parents to discover what happened, they would likely blame me for my negligence, but they should not react as severely as your friends react toward you.

Cases like these evoke reactions that are difficult to reconcile. Both you and I were negligent. We were negligent in precisely the same way. The fact that your negligence led to tragedy and mine didn't was entirely out of our control. Good or bad luck, it seems, cannot affect our levels of moral responsibility—cannot influence the amount of guilt we should feel, nor the amount of blame we deserve.

Like so many problems in philosophy, the problem of moral luck arises out of a clash between a compelling abstract principle and intuitions about particular cases. As a general principle, it seems, people are responsible only for what is within their control. However, we frequently, consistently, and plausibly hold people responsible on the basis of factors that are beyond their control. We balk at moral luck in the abstract, but our moral practices are suffused with it.

I will argue that moral luck is real and not illusory: people are genuinely responsible for some things that are beyond their control. I will begin by criticizing skepticism about moral luck. I will then make a positive case for the existence of moral luck by appealing to a mixed theory of moral responsibility. Although blame and punishment have a retributive rationale, I contend that they also have a consequentialist rationale. Responsibility cannot be leashed to control if it is to meet the consequentialist portion of its rationale. Blame and punishment are therefore partly a matter of luck.

A central contribution of this essay is that recent empirical work supports the existence of moral luck. Studies of blame, punishment, and learning shed new light on our practices of holding one another morally responsible. In particular, the work suggests that our luck-sensitive moral practices are historically rooted in the very same consequentialist considerations that justify moral luck. In short, there are good reasons for responsibility to outstrip control, and these reasons are not mere rationalizations of the status quo: they have played a role in shaping our sensitivity to luck.

1. Skepticism

Skepticism about moral luck is the view that responsibility is never a matter of luck—that good or bad fortune cannot affect one's level of responsibility (Smith 1759; Richards 1986; Thomson 1993; Rosebury 1995; Wolf 2001; Zimmerman 2002). This view rests on an appealing principle often called the "control condition": people are responsible only for what is under their control. However, as we'll now see in detail, cases of apparent moral luck are intuitively compelling and difficult to deny.

There are several types of moral luck, each of which presents a unique challenge to skepticism. Consider first what Thomas Nagel (1979) calls "resultant" moral luck. Two agents perform the same action but seem to bear different amounts of responsibility because of further, downstream effects that are beyond their control. Resultant moral luck arises in cases of negligence, like those of the lucky and unlucky drivers. But it also arises in cases of intentional wrongdoing. Suppose that two people both shoot with the intent to kill. The first succeeds, whereas the second person's gun jams. The first person has caused the death of another and thus, it seems, is more blameworthy than the second person. The first is guilty of murder, the second only of attempted murder.

Skeptics hold that we should treat cases alike if the degree of control in each case is the same. Some might suggest that the best way to do this in difficult cases is to balance responsibility toward an average, i.e., decrease responsibility for unlucky agents and increase responsibility for lucky agents. But this seems to yield two implausible verdicts instead of only one. Should we inflict less punishment on a person who drives drunk and kills an innocent person, sending him to jail for a few months instead of several years? Send a lucky drunk driver to jail for a few months instead of fining him and revoking his license? Balancing responsibility toward an average seems no more plausible than assimilating one case to the other (Wolf 2001: 7).

Skeptics argue that our intuitions about cases of resultant moral luck are mistaken, perhaps distorted by cognitive biases, and should be abandoned. In the face of a clash between an abstract principle and intuitions about cases, one option is to abandon the principle. But, of course, another option—as skeptics urge—is to override potentially wayward intuitions. The problem, however, is that there are other types of moral luck, and they present even more difficult challenges for skepticism—intuitions that are even more difficult to override.

In what Nagel calls "circumstantial" moral luck, two people who perform different actions seem morally responsible to different degrees even though their actions were different solely in virtue of circumstances utterly beyond their control. Had those circumstances been the same, they would have performed the same action. Thus, imagine someone who is guilty of a crime of passion. She assaults her cheating partner. Another person, however, is innocent of this crime simply for luckily not having been provoked. Had she found out that her partner was cheating, she would have reacted with the same violence. And yet, the first person seems blameworthy to a degree that the second person isn't. Or consider, as Nagel recounts, the many citizens of Nazi Germany who committed or supported abhorrent crimes against Jews and other vulnerable minorities for which they must bear responsibility, even though citizens in many other countries almost certainly would have done so at the same rate under the same social and political conditions. Stanley Milgram's famous obedience experiments, in which subjects agree to inflict painful and even purportedly lethal electric shocks on innocent people, show that circumstances can radically alter agents' propensity to engage in immoral behavior.

According to skeptics, we should hold otherwise similar people responsible to the same degree, ignoring all effects upon their actions due to circumstance. Intuitively, however, this seems unacceptable. We shouldn't punish two people equally for a crime of passion that only one of them has committed, notwithstanding that the other would have committed the same wrong had she been placed in similarly unlucky circumstances. Skeptics must recommend a level of blame and punishment that is appropriate to inflict upon criminals as well as those who would have been criminals in the same circumstances but in fact did nothing. However, while some of us have been lucky by escaping an unfortunate outcome arising from our action, others have been circumstantially lucky by not performing any culpable action at all.

The problem of "constitutive" moral luck is even more difficult for skeptics. Our character traits, abilities, and motivations are the product of genetic and environmental conditions that are almost unquestionably beyond our control. One person is kind, another is cruel; one person is courageous, another is cowardly. But which trait one possesses is the result of luck. Admittedly, actions have the power to shape one's character. But unless we are "unmoved movers," constitution-shaping actions are themselves the product of prior genetic and environmental conditions. So, to bracket off from moral responsibility any aspects of a person's character, abilities, or motivation that are outside of her control would be to eliminate everything. We would be forced to blame and punish everyone to the same degree.

Skepticism about moral luck is tempting. It seems unfair or unreasonable to blame a person for something that she cannot do anything about. This is the thought that underlies the control condition. From the control condition it follows that responsibility is leashed to control: all else being equal, two people cannot differ in their level of responsibility without differing in their level of control. Thus, skeptics admit that the control condition conflicts with our intuitions and practices but argue that because the principle is so compelling we should override our intuitions and reform our practices. However, as we've seen, denying moral luck is intuitively unacceptable. With respect to some types of moral luck, denial may be simply impossible. Skepticism is therefore untenable because it requires drastic, even unattainable revisions to intuitively appropriate moral practices.

My attention in the rest of the essay will be exclusively with resultant moral luck. This type of moral luck arises in cases of intentional wrongdoing and negligence. In both cases, the outcome of an agent's action or inaction seems to influence the degree of blame and punishment that she merits. To evaluate whether or not resultant moral luck is real, it will help to begin by understanding the psychological dispositions that underlie assignments of blame and punishment. With this empirical understanding in hand, we will be in a better position to assess the distinctively normative phenomenon of moral luck. My contention in the essay, more broadly, is that empirical work in psychology and evolutionary theory, in combination with philosophical considerations, can help us understand whether moral responsibility outstrips control.

2. The Psychology of Moral Luck

Research by Fiery Cushman and other scientists seems to explain why, in our practices, we are sensitive to resultant luck. As I will show by the end of this section, this research reveals the extent to which the control condition is, and is not, consistent with those practices. In the next section, I will also draw on Cushman's work as I consider a prominent defence of skepticism about moral luck. Whereas the skeptics considered above recommend reform of our practices, other skeptics argue that the control condition does not in fact conflict with our practices. Proponents of the "epistemic view" begin by noticing that the outcomes of our actions are often the best evidence of our intentions. So, they argue, we blame and punish people differently depending on the varying degrees of harm their actions have brought about because this is evidence, albeit defeasible, that they acted on different intentions. Ultimately, I will argue that the epistemic view is implausible.

Let's begin, however, with a clearer look at our practices of assigning blame and punishment. Cushman's (2008) earliest research on this topic provides an important clarification about our sensitivity to resultant luck. His research uncovers an asymmetry: between judgments about blame and punishment relative to other moral judgments. Participants in Cushman's initial study were presented with cases of intentional wrongdoing and cases of negligence, identical except that harm either does or doesn't occur as a result. They were then asked (1) whether the agent's action is wrong, (2) whether it is impermissible, (3) whether she is blameworthy, and (4) whether she should be punished. It turns out that judgments about whether an agent's action is wrong or impermissible depend almost exclusively on her mental states, i.e., whether she intends harm or whether she negligently failed to form an intention to avoid foreseeable harm. By contrast, judgments about whether an agent should be blamed or punished for her action depend both on her mental states *and* the harmful outcomes of her actions. When an agent commits an intentional wrong or is negligent, the outcomes of her action or omission influence ascriptions of responsibility, but they do not influence ascriptions of wrongness or impermissibility.

In follow up work, Cushman et al. (2013) probed another type of moral judgment. Instead of simply asking whether an agent's action is wrong or impermissible, or whether the agent should be blamed or punished, the researchers also asked participants about the agent's moral character. Like wrongness and impermissibility judgments, character judgments are influenced primarily by the agent's mental states and not appreciably by the outcomes of her action. Thus, overall, participants judged an agent to have poor moral character if she intended harm and good moral character if she didn't intend harm, largely irrespective of the harm that the agent actually caused.

These empirical findings, Cushman argues, suggest that our judgments about blame and punishment reflect the influence of two different processes of moral evaluation (Cushman 2008; Cushman et al. 2013). One process is activated by harmful outcomes, analyzes the action that caused the harm, and then assigns blame and punishment to the causally responsible agent. The other process analyzes an agent's mental states, determines whether she intended harm or was negligent, and then not only assigns blame and punishment to the agent, but also yields judgments about the wrongness and impermissibility of her actions and about her character. The total amount of blame and punishment assigned depends on the contribution of both processes. That is, both outcomes and mental states influence ascriptions of responsibility. However, harmful outcomes do not affect ascriptions of responsibility when no culpable mental state is present (at least not typically). Thus, if an agent acts rightly, but an unforeseeable harmful outcome nonetheless occurs, blame and punishment are typically not assigned.

Cushman's most important finding is that there is an asymmetry between judgments of wrongness, impermissibility, and character relative to judgments of blame and punishment. The former depend almost exclusively on the agent's mental states, while the latter depend both on her mental states and the outcomes of her action. Thus, only judgments of blame and punishment depend on luck. To be clear, this empirical fact about human psychology does not by itself lend any support to a philosophical view about moral luck. But it does shed light on the phenomenon at issue. If we want to make sense of our moral practices, we must know precisely what they are.

To see this, notice that some authors seem to conflate responsibility judgments with other moral judgments. Judith Thomson (1993) defends skepticism about moral luck by appealing to intuitions about how neither wrongful action, nor the character of the agent who performed an action, could possibly depend on luck. These intuitions are indeed compelling. However, those who wish to defend the existence of moral luck, if they have an eye on our actual practices, should respond by claiming that lucky or unlucky outcomes seem to bear only on blame and punishment, and that they do not impact wrongness, impermissibility, or character (Nelkin 2013).

Cushman's empirical findings about the psychological processes underlying our responsibility practices do not have immediate philosophical significance. However, they do shed light on exactly what it is that defenders of moral luck should seek to preserve. Thus, Cushman's findings suggest the possibility of a *restricted* version of the control condition that is consistent with our practices. Application of the control condition to wrongness, impermissibility, and character is perfectly consistent with our practices. An action is wrong or impermissible, and reflects well or poorly on the character of the agent who performs it, indeed, only insofar as the agent has control over what she does. Thus, even if they should receive different amounts of blame and punishment, two people who neglect to secure a child in their vehicles—no matter what the outcome—perform actions that are equally wrong, that are equally impermissible, and that reflect similarly on their moral characters. Intuition does not rebel at this application of the control condition.

Furthermore, we do apply the control condition to blame and punishment, but only to one determinant of blame and punishment. Someone is fully responsible for a harmful action if he intends harm and actually brings it about. But he is less than fully responsible if he merely attempts harm (he intends harm but it doesn't occur) or if he happens to avoid foreseeable harm (he was negligent but harm doesn't occur). Thus, our intuitions and practices do not support "strict liability" about moral responsibility; outcomes are not the only thing that matters, and they matter only when an agent intends ill or is negligent. As it plays out in our practices, moreover, the control condition is a constraint on only one of the two "vectors" of responsibility. It constrains the mental state vector but not the outcome vector.

The control condition is compelling because some version of it serves as a constraint on our moral judgments. It appears, however, that we treat the control condition as a full constraint on wrongness, impermissibility, and character, and only as a partial constraint on blame and punishment. So, one can affirm the importance of the control condition without being a skeptic about moral luck—that is, by accepting a restricted version of that principle. For now, I am simply describing the influence of control on our moral judgments and the way in which our judgments are compatible with a restricted version of the control condition. This softens the blow of moral luck upon an appealing general principle of moral responsibility. Later on, I will explore whether anything positive can be said in favor of the existence of moral luck. My next task, however, is to examine and criticize what may be the most important line of defence for skepticism.

3. The Epistemic View

As we've seen, some skeptics reject and attempt to override intuitions that seem to support the existence of moral luck (Thomson 1993). Other skeptics, however, seek to explain away apparent cases of moral luck as epistemic phenomena (Richards 1986). Proponents of *the epistemic view* argue that blame and punishment are assigned on the basis of intentions and outcomes, rather than intentions alone, for the reason that intentions are often somewhat opaque. Thus, outcomes are often an important source of evidence about what one intends. If someone successfully brings about harm, for example, this is good evidence that she intended to do so. And if someone causes a tragic outcome, this is good evidence that she was egregiously negligent.

In an early and especially clear articulation of the epistemic view, Norvin Richards (1986) distinguishes between the blame and punishment that someone deserves and the blame and punishment that we are entitled to assign to her on the basis of our evidence. Luck, Richards argues, does not affect the former but it does affect the latter. Thus, downstream events beyond an agent's control influence not her responsibility itself, but only the epistemic position of those who must hold her responsible. Some people are lucky or unlucky with respect to their justified exposure to sanction, without being lucky or unlucky with respect to the amount of sanction they truly deserve.

According to the epistemic view of moral luck, outcomes do not matter for responsibility in themselves, but they are a reliable source of evidence about what does matter, i.e., one's mental states. So, because treating outcomes as evidence is consistent with an *un*restricted version of the control condition, the tension between this abstract principle about responsibility and our practices is, allegedly, only apparent and not genuine. The problem with skepticism, at first glance, is that it requires drastic and implausible changes to the way we assign moral responsibility. But, on the epistemic view, so it would seem, no such radical reform of our practices is necessary.

The epistemic view is problematic, however. It ultimately fails to make sense of our practices. I will develop two separate criticisms, each of which is supported by a combination of empirical and philosophical considerations. The first criticism is rooted in the asymmetry laid bare by Cushman and described in the last section. When people make judgments about the moral wrongness or impermissibility of an action, or about the moral character of the agent who performed the action, they rely exclusively on intentions. But when people make judgments about the blame and punishment that an action deserves, they rely on intentions and outcomes. The epistemic view says that we treat outcomes as evidence of an agent's mental states. However, the view then predicts that outcomes should *also* influence judgments about wrongness, impermissibility, and character. For here too, by hypothesis, outcomes should serve as reliable evidence about an agent's intentions, and intentions are what matter for wrongness, impermissibility, and character. Cushman's studies contradict this prediction, as we've seen, and therefore suggest that outcomes are not generally treated merely as a proxy for intentions. If we *were* to treat outcomes as a proxy for intentions, then another massive and morally implausible overhaul of our practices is necessary, not to our judgments about blame and punishment, but to our judgments about wrongness, impermissibility, and character. That is, we would have to rely on outcomes in assessments of wrongness, impermissibility, and character, too. We would be forced to judge that when your negligence leads to a child's injury, and my negligence doesn't, that your action is wrong to a greater degree than mine. Thus, skeptics cannot escape radical and unappealing revision.

A second criticism of the epistemic view is that even when we are certain of agents' intentions, it still seems appropriate to hold them responsible to different degrees based on the outcomes of their actions. For example, even if we know without a doubt that two agents both intended to shoot someone, the fact that one hit and the other missed seems to warrant different amounts of blame and punishment. This is, perhaps, even clearer in first personal cases. I do not need to observe outcomes to determine my own intentions, at least not typically, and yet whether or not my action resulted in harm influences the amount of blame I impose on myself. The epistemic view predicts that outcomes are only indirectly relevant to responsibility, but we consistently treat them as directly relevant.

A study by Cushman et al. (2009) yields experimental confirmation that people generally rely on outcomes to assign responsibility even when they are certain of agents' intentions. In this study, pairs of participants are recruited for a game designed to involve luck. One participant, Giver, is presented with a sum of money and instructed to offer none, half, or all of it to Receiver. Thus, Giver can be either stingy, fair, or generous. However, Giver is instructed to execute her choice by rolling one of three dice. A stingy die is likely to give all the money to herself; a fair die is likely to split the money equally; a generous die is likely to give all of the money to Receiver. But, as both Giver and Receiver know, rolling any of these dice carries the probability of an unintended outcome. For example, if Giver chooses the stingy die, there is a 4/6 chance that Receiver will get nothing, as she intended, but there is a 1/6 chance that the split will be equal, and a 1/6 chance that Receiver will get all the money.

The next phase of the experiment is critical. Receiver is given a chance to reward or punish Giver by adding money to, or subtracting money from, her total payoff. Critically, Receiver is made aware of not just how much money he received but also which die Giver chose, and thus her intentions. If Receiver cared only about intentions, we should expect him to punish Giver whenever she chose the stingy die, irrespective of how much money he actually received. Conversely, if Receiver cared only about outcomes, we should expect him to punish Giver whenever he received no money, irrespective of which die she chose. What Cushman finds is that participants in Receiver's role tend to punish on the basis of intentions *and* outcomes. So, for example, when Giver intended to be stingy, Receiver tended to punish her more when the outcome was stingy than when it was fair or generous. Even when people are certain of others' intentions, then, they continue to hold them responsible for outcomes. Outcomes are not treated merely as an epistemic proxy for intentions.

Proponents of the epistemic view wish to reconcile our moral practices with an unrestricted version of the control condition and thus explain away the appearance of moral luck. They rely on an empirical hypothesis. They argue that when we assign blame and punishment to others, we rely on outcomes only as evidence of intentions. However, our practices don't seem to bear this out. No reconciliation, it seems, is in the offing. And so, skepticism is in fact saddled with radically revisionary implications that proponents of the epistemic view seek to stave off.

4. Debunking Moral Luck

As we've seen, intuitions about several different varieties of moral luck challenge the control condition. Furthermore, so far as we can tell thus far, intuitions about resultant moral luck cannot be explained away as epistemic phenomena, and thereby rendered consistent with the control condition. Yet, skeptics have one more line of defense: to debunk the intuitions they find problematic by exposing allegedly defective psychological and evolutionary processes underlying them. These debunking arguments come in two flavors, corresponding to the two main skeptical strategies. One relies on the epistemic view and attempts to save our intuitions while deflating their significance. The other debunking argument claims that we should abandon luck-sensitive intuitions.

Let's begin with Neil Levy (2016), who offers an empirically-grounded defense of the epistemic view. In a study by Heather Lench et al. (2015), researchers found that outcomes influence ascriptions of blame and punishment, replicating Cushman's results. But Lench and colleagues also conducted a mediation analysis, the purpose of which was to examine causal relationships between different elements in the underlying psychological mechanism. Their mediation analysis suggests that harmful outcomes cause ascriptions of negative intentions, which then cause assignment of blame and punishment (see also Young et al. 2010). This result is good news for proponents of the epistemic view, Levy argues, since it suggests that we treat outcomes as evidence of intentions.

The difficulty, however, is that on Lench et al.'s mediation analysis, there remains a direct effect of outcomes on blame and punishment. So, even if outcomes are treated as indirectly relevant to responsibility they are also treated as directly relevant. Outcomes might be an epistemic proxy, but they are not only that. Contrary to Levy, then, Lench et al.'s mediation analysis does not rescue the epistemic view.

Levy argues next, however, that the psychological mechanism that underlies blame and punishment is "encapsulated." That is, the mechanism is isolated from explicit beliefs and from conscious reasoning. So, Levy argues, this is why even when we recognize explicitly that an agent's intentions are more benign than the outcomes of her action suggest, we remain disposed to blame and punish her. He suggests that evolution has designed us to inflexibly treat outcomes as reasons for assigning responsibility because outcomes are statistically correlated with intentions. Thus, we unconsciously employ an "outcome heuristic" that is a reliable shortcut for inferring responsibility. In that case, we shouldn't expect the effect of outcomes on ascriptions of responsibility to be entirely mediated by intentions. Moreover, because of our built-in moral blindness, we shouldn't trust intuitions that support the existence of moral luck.

Levy's psychological and evolutionary hypotheses confront two main problems. The first is similar to the problem that afflicts other proponents of the epistemic view and rests on the asymmetry that is by now quite familiar. If we are designed to inflexibly treat outcomes as reasons for blame and punishment because of the statistical correlation between outcomes and intentions, then we should expect outcomes to affect judgments of wrongness, impermissibility, and character, too. That is, Levy's evolutionary argument predicts that the mechanism underlying these other moral judgments would be similarly encapsulated, such that they inflexibly treat outcomes as reasons for moral judgment about the wrongness and impermissibility of an agent's actions and her character. In both cases, evolution could have taken advantage of the regular connection between outcomes and intention. However, Cushman's asymmetry shows that the two mechanisms are not similarly designed. It's not clear why the mechanism underlying action and character judgments isn't similarly encapsulated. Levy is aware of this problem, but his evolutionary hypothesis does not explain why evolution failed to design us to inflexibly treat outcomes as reasons for moral judgment across the board.

The second problem is that the psychological mechanisms underlying blame and punishment seem *not* to be encapsulated. When we reassess someone's intentions, and realize they were benign rather than malign, we often reassess our attribution of responsibility. For example, if someone slaps your friend's shoulder and causes her evident pain, you might initially interpret his act as aggressive and immediately feel disposed toward blame. If you realize, however, that he was swatting away a bee and someone else in your party is allergic to bees, the blame fades away. Unlucky bad outcomes matter only when the agent engages in intentional wrongdoing or is negligent. The psychological mechanism underlying ascriptions of responsibility is not encapsulated because when we revise our beliefs about whether someone acted wrongly or was negligent our ascriptions of blame then typically change too. In response to counter-examples like these, which are plentiful, Levy might reply that beliefs only provide new inputs to a process that is otherwise encapsulated. But then we should expect beliefs about the control condition to provide inputs to blame and punishment in cases of apparent resultant moral luck. We don't find this, of course: our intuitions do not disappear when we explicitly face the problem of moral luck.

Levy suggests that intuitions about moral luck reflect an underlying rationale: they track intentions, either as a matter of cognitive or evolutionary design. Certain authors likewise think that the intuitions should not be taken to support the existence of moral luck. However, they argue that the intuitions, rather than products of well-designed if occasionally faulty heuristics, are subject to cognitive biases. Two sets of authors pursue this type of debunking argument. Like the first set of skeptics we confronted, in section one, these debunkers argue that in the face of a conflict between the control condition and intuitions about cases, we should abandon intuitions.

Edward Royzman and Rahul Kumar (2004) argue that intuitions are distorted by the bias of hindsight. That is, people tend to irrationally infer from a harmful outcome that an agent should have known that her action was likely to be harmful. Thus, when no harmful outcome occurs, people infer the agent was slightly negligent; when a harmful outcome does occur, people infer that the agent was extremely negligent. So, the hindsight bias explains—and debunks—our sensitivity to luck. Darren Domsky (2004) also argues that the intuitions are distorted by bias, but he identifies selfishness and irrational optimism as the culprits. A person is more likely to blame others when their actions yield harmful outcomes because she is disposed to believe that she would have been able to act with greater care. Thus, Domsky argues, these biases trick us into thinking that morality is subject into luck.

The debunking arguments developed by these authors rely on a rich set of evidence about general biases that afflict judgment and decision making. However, if Cushman's account is right, intuitions about moral luck issue from two dedicated mechanisms for assigning responsibility. No general biases are at work, evidently, since they don't seem to have the same influence on judgments about wrongness and impermissibility. If hindsight bias led people to infer bad intentions from bad outcomes, luck would influence all of our moral judgments similarly, not just those that concern blame and punishment. Once again, an argument for skepticism founders on Cushman's asymmetry. The two sets of debunkers might be tempted to suggest that the biases are specific to a single mechanism (and not the other mechanism). However, this suggestion would be ad hoc; the debunking arguments would lose the appeal that their hypotheses otherwise gain from independent evidence that these biases affect so many other judgments and decisions.

So far in the essay, I have mounted a philosophical and empirical case against skepticism about moral luck. Skepticism conflicts with powerful intuitions about responsibility that underlie widespread and seemingly appropriate practices of holding one another responsible. Those who merely attempt harm and those who successfully bring it about seem, intuitively, to be responsible to different degrees. Likewise for those whose negligence leads to harm and those whose negligence doesn't. As we have just seen, these intuitions are not simply the product of faulty cognitive heuristics or biases. Skeptics appeal to an admittedly attractive control condition on moral responsibility, but a restricted version of that condition is available that applies fully to wrongness, impermissibility, and character, and still partially to blame and punishment. That is, a restricted version of this attractive principle is fully compatible with the existence of moral luck.

Some authors attempt to explain away the appearance of moral luck in our practices as an epistemic artifact, and thereby save skepticism from its drastic revisionary implications. In fact, however, we do not treat outcomes as evidence of intentions when it comes to wrongness and impermissibility, and we consistently hold people responsible for outcomes even holding fixed our knowledge about their intentions. Our dispositions to blame and punish others are genuinely sensitive to lucky and unlucky outcomes. It's now time to determine whether any sense can be made of our sensitivity to moral luck.

5. Retributivism & Consequentialism

In the rest of the essay, I will offer a series of arguments for realism about moral luck (cf. Walker 1991; Moore 1997, 2009; Woodrow 2006). If we let go of skepticism, we must abandon an unrestricted formulation of the control condition. But, in that case, what general theory or principle concerning moral responsibility, if not the control condition, should we accept? That is, what moral doctrine explains why we should blame and punish people on the basis of lucky or unlucky outcomes? In this section, I lay out a general philosophical theory of moral responsibility and show how it supports the existence of resultant moral luck. In the following sections, I will respond to objections and go on to defend a more specific formulation of the general theory, drawing once again on a combination of empirical and philosophical considerations.

Responsibility is, I suggest, partially retributive, but also partially consequentialist. To be a retributivist is to think that responsibility is a matter of praising and blaming people in accordance with what they merit—the most widely endorsed candidate for conferring merit is their intentions (or their will or their character). But one can accept a retributivist view and yet think that responsibility is *also* a matter of bringing about good consequences. The central consequentialist justification for blame and punishment is deterrence, i.e., preventing people from behaving immorally in the future. On a mixed view, then, we should blame and punish people partly because they deserve censure and also partly to deter future misbehavior.

A mixed view of moral responsibility is plausible because it seems to get the best of both worlds: the rationale for blame and punishment encompasses both what is under the agent's control and what will improve people's lives. But how does a mixed view bear on moral luck? Why, in light of this view, should we blame and punish people partly on the basis of the outcomes of their actions and not just on the basis of their intentions? The answer, I believe, is that this facilitates deterrence of immoral behavior. This is why a mixed view of responsibility makes sense of resultant moral luck. After explaining this idea in detail, I will spend the rest of the essay defending it.

At first blush, it may seem as if punishing those who have malicious intentions or who act out of negligence is the best way to secure positive consequences—malicious and negligent people present the greatest threat to others and thus it is their behavior that most begs for deterrence. However, as proponents of the epistemic view notice, outcomes are typically much easier to identify than intentions. It is usually obvious whether or not harm occurred. But it is often difficult to know whether someone genuinely intended harm or made a mistake, difficult to know whether they were negligent or caused harm despite exercising due care. So, blaming and punishing people strictly for their intentions would be unreliable: we would hold many people responsible who we mistakenly classify as ill-willed or negligent; we would also fail to hold many people responsible whose ill-will or negligence is not apparent. 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.

The point here is not what proponents of the epistemic view emphasize, i.e., that outcomes track intentions. Rather, it's that norms that link responsibility to outcomes can be applied with greater accuracy, compared with norms that link responsibility solely to intentions. We can more easily identify outcomes than intentions. Thus, a community that employs outcome norms is more likely to accurately apply *its own standards*. Relatedly, because outcomes are publically accessible, outcome norms are more likely to secure third-party agreement, and thus facilitate social coordination of blame and punishment. Furthermore, suppose that Peter Carruthers (2011) is correct that what seems like introspection is in fact first personal theory of mind. Attribution of mental states to ourselves depends on the same behavioral and linguistic evidence that we use to attribute mental states to others. In that case, our own intentions are opaque too. On grounds of deterrence, then, we should not attempt to blame and punish others—or perhaps even ourselves—solely on the basis of intentions.

To be clear, although one purpose of holding others responsible is deterrence, a mixed view of responsibility entails that blame and punishment have other, non-consequentialist purposes, too. A partially consequentialist theory of blame and punishment is, for example, consistent with the idea that another role for blame is to assess the quality of a person's will or to register a change in the status of a relationship (see Scanlon 2008).

A mixed view, I have argued, offers an attractive account of not just blame and punishment generally, but also resultant moral luck in particular. That is, it makes sense of the practices that Cushman's research describes. We blame and punish based on intentions, which makes our practices justified on traditional retributive grounds. But we also blame and punish based on outcomes, and since this makes our practices more reliable and thus a more effective method of deterrence, our luck-sensitive practices also satisfy a consequentialist rationale. This is why a partial—but only partial—commitment to consequentialism favors moral luck.

Now that I have set out a general theory of moral responsibility, one that supports the existence of moral luck, I want to spend the next section addressing objections. I will offer answers to these objections, but empirical work discussed later on will provide further resources for answering them. Furthermore, the objections will lead us to one type of mixed view that seems to be more plausible than others.

6. Objections

First of all, skeptics might object that consequentialism cannot genuinely provide an account of moral responsibility per se. Responsibility is a matter of the blame and punishment that someone *deserves* or *merits*. Pressing this point, the critic might claim that consequences can provide only *incentives* to engage in blaming and punishing behavior. To put this in terms that are current in metaethics, deterrence provides only the "wrong kinds of reasons" to blame and punish; it doesn't provide the "right kinds of reasons" related to desert or merit. This is an important objection. If deterrence provides only the wrong kinds of reasons, then I should reformulate the positive claims I have been making. I should say not that moral responsibility is genuinely a matter of luck, but that we have good (practical) reasons to blame and punish others as if it were.

However, another possibility is that indirect consequentialism can provide an account of normative categories like desert and merit (cf. MacKenzie 2017). This form of indirect consequentialism is a "two-level" theory of the sort suggested by John Rawls (1955) and H. L. A. Hart (1968), among others. Essential to a two-level theory is the separation of two questions, one about the justification for our acts or attitudes and the other about the justification for the norms or practices that underlie our acts and attitudes. According to a two-level theory, retributivism answers the first question, while consequentialism answers the second. First, what justifies us in holding someone responsible? Her intentions and the outcomes of her action; these are the basis upon which punishment is deserved or merited. Second, what justifies the practice of holding someone responsible according to these norms that advert to intentions and outcomes (rather than according to other, incompatible norms)? The answer to this question is that our practices bring about good consequences. Our practices of blame and punishment are justified because they facilitate deterrence. At a first order level, intentions and outcomes determine what a person deserves or merits. But at a second order level, these norms of desert or merit have a consequentialist rationale. In the context of moral luck, in particular, norms that enjoin blame and punishment based partly on intentions and partly on outcomes provide an effective means of deterring immoral behavior.

The first objection at hand claims that consequentialism is unable to provide an account of normative categories like merit or desert. However, indirect consequentialism is better suited to meet this challenge than direct consequentialism (where consequences justify not acts and attitudes themselves, but the norms, rules, codes, or practices underlying acts and attitudes). The idea is that good consequences generate reasons to adopt norms of desert or merit. We have second order consequentialist reasons, that is, to accept these first order, *non*-consequentialist norms. The first order norms do not consider consequences and concern only what is deserved or merited.

I do not claim to have argued for indirect consequentialism about moral responsibility over all competing views. In particular, indirect consequentialism faces general challenges that must be addressed if it is to be regarded as plausible in any ethical domain. For example, critics argue that indirect consequentialism has a problem with "rule worship" or that indirect consequentialism reduces to direct consequentialism (see, e.g., Hooker 2016 for discussion). I do not have the space to engage with these rather large issues here. However, I have argued that a mixed view of moral responsibility is in a unique position to make sense of resultant moral luck, and that indirect consequentialism is the most plausible version of a mixed view, in light of the foregoing objection. Furthermore, as we'll see, the arguments developed in the next section support an indirect consequentialist rationale for moral luck.

Let's turn now to a second and, I believe, equally important objection. Skeptics might argue that a consequentialist argument for moral luck is merely a post-hoc rationalization of our moral practices, and thus an untrustworthy application of confirmation bias. Perhaps whatever our practices had been like, we would have been able to find some consequentialist justification for them. If so, then we should not be confident that our current practices genuinely have good consequences in the way that I have suggested. In the same way that evolutionary "just-so" stories can provide an adaptive rationale for any trait, even when it isn't genuinely an adaptation, consequentialist justifications threaten to be "just-why" stories (Kumar 2017; Kumar and Campbell 2018). Their ubiquitous availability casts doubt on any particular application.

Consequentialism is especially susceptible to the problem of post-hoc rationalization (contra, e.g., Greene 2008). To illustrate, consider a well known problem for consequentialism: that it seems to conflict with the idea that we have special duties to friends, family members, and other intimates. Why should you pay for your own child's education when the same money might do more good for another, more capable child who can't afford to attend college on her own? A common reply one hears from consequentialists is that you are in a better position to help your children than you are to help others. Often, this reply strains credulity. In many cases, charity toward strangers is not especially difficult. But, moreover, the fact that a plausible-sounding consequentialist story always seems available to justify the status quo, *whatever the status quo happens to be*, suggests that consequentialist justifications are unreliable and casts doubt on them. And so, it seems, this fact cast doubt too on a consequentialist justification for resultant moral luck.

I will argue, however, not just that there are consequentialist reasons for resultant moral luck, but that our practices are shaped by these reasons. That is, our practices of blame and punishment are sensitive to outcomes *because* this sensitivity yields the positive consequences that justify it. The consequentialist rationale is therefore not just a post-hoc rationalization; it's reflected in the etiology of our sensitivity to moral luck. The burden of the next section is to argue that our practices are *based on* the consequentialist considerations that justify our practices.

Notice first, however, that the idea that our sensitivity to luck is based on consequentialist considerations confronts an apparent difficulty. Blame and punishment are typically offered when they are thought to be merited, not when they are thought to be expedient. Studies consistently find that people are reluctant to let consequentialist considerations influence their decisions to blame and punish. For example, researchers present participants with cases in which punishment is either more or less likely to have good consequences. These differences tend not to change participants' judgments about how much punishment is appropriate (Carlsmith 2006, 2008). Furthermore, in economic games that are one-off and anonymous, participants punish partners who play unfairly (Fehr and Gachter 2002), even when they can expect their partners to be unaware that they have been punished (Nadelhoffer et al. 2013), suggesting that punishment is not motivated by a motive to change their partner's behavior.

People tend not to blame and punish others with the objective of deterrence. In what sense, then, are our practices of blame and punishment based on consequentialist considerations? As I will explain, the basis is indirect: consequentialist considerations have shaped the evolution of our responsibility practices. Attending directly to outcomes led to better methods of deterrence, and this consequence shaped the evolution of the practice. We subscribe to certain non-consequentialist norms of blame and punishment, but we subscribe to these non-consequentialist norms ultimately for consequentialist reasons. Thus, if we look at the way in which our norms of moral responsibility arose, a two-level theory of moral responsibility—a form of indirect consequentialism—is reflected in their etiology.

7. Evolution & Learning

In this section, I will explore the evolution of moral responsibility. I will begin by briefly discussing the central puzzle for evolutionary theorizing about morality. The roots of morality lie in altruism and cooperation. How did these forms of pro-sociality evolve? If pro-sociality is, at some level, an adaptation for social creatures like us, how was it selected for? Of course, pro-social behavior often benefits others at a net cost to oneself. So, how could pro-sociality evolve when it decreases individual fitness?

A number of different solutions to this puzzle have been proposed. However, it is likely that punishment offers the most important key to resolving the puzzle (Boyd and Richerson 1992). Broadly, when anti-social behavior is punished, pro-social behavior confers a net fitness advantage on those who engage in it. This is one of the best current theories that explains how a basic disposition toward pro-sociality evolved (for more detail see, e.g., Kumar and Campbell 2018). Furthermore, then, the evolution of punishment embodies a consequentialist rationale: punishment deters anti-sociality and thus secures the group level benefits of pro-sociality.

Even if the foregoing account is part of an accurate explanation of the co-evolution of punishment and pro-sociality, and even if in the process it reflects a partially consequentialist rationale for punishment, it does not explain, more specifically, how we evolved to assign responsibility partly on the basis of the outcomes of people's actions. Why are we sensitive to luck? We seem to know the proximal causes in our psychology, from Cushman's experimental work, but what is the ultimate cause? The answer to this question, according to an evolutionary model developed by Cushman, lies in the dynamics of punishment and learning (Cushman 2013, 2015; Martin and Cushman 2016). I will discuss Cushman's evolutionary model in this section, but my aim is to draw philosophical conclusions that go beyond his empirical work.

To understand Cushman's view, we must begin by noting that pro-sociality is flexible, whereas punishment is rigid. Although we likely evolved an innate pre-disposition toward pro-social behavior, Cushman argues that learning mechanisms also facilitate pro-sociality. In circumstances where pro-sociality pays, we learn to become more pro-social and less selfish. But in circumstances where pro-sociality is costly, we learn to become less pro-social and more selfish. Punishment is a mechanism that encourages people to be pro-social, by introducing disincentives for selfishness that we are equipped to track and respond to. Cushman points out, however, that whereas pro-sociality is flexibly shaped by incentives and disincentives, dispositions to punish are more rigid, that is, relatively insensitive to incentives and disincentives. As reviewed at the end of the last section, people are disposed to punish others whether or not it benefits them or anyone else.

According to Cushman, we evolved to punish rigidly based on the outcomes of other people's actions because this was necessary to secure the social learning conditions that favor pro-sociality. A rigid disposition to punish for outcomes was thus selected for over alternative punishment strategies. There are two parts to this argument.

On the one hand, punishment based on outcomes has an advantage over punishment based only on intentions. The reason is that those who punished based only on intentions allowed anti-social agents to manipulate them into believing that their intentions were benign. This is just the opacity of intentions all over again, but in a strategic context. Punishing partly on the basis of outcomes, by contrast, prevents manipulation; it is, of course, difficult and usually impossible to convince someone who has been harmed, or has witnessed it, that the harm did not occur. So, this is why outcome-based punishment is superior to pure intention-based punishment.

On the other hand, rigid punishment has an advantage over flexible punishment. Those who punished with attention to the consequences of their punishment (i.e., flexibly) allowed anti-social actors to disincentivize punishment, for example, by exhibiting and communicating an unwillingness to learn from punishment. If your motive for punishing an offender is to reform him, you will be unmotivated to punish him if you are made to believe that reform is impossible. Incorrigible anti-sociality would then have had a fitness advantage over pro-sociality. So, this is why rigid punishment is superior to flexible, deterrence-motivated punishment.

Cushman concludes that only rigid, outcome-based punishment was able to provide a learning environment that favors pro-sociality over anti-sociality. Punishment that is either affixed solely to intentions or motivated by deterrence is inferior. Evolutionary dynamics thus explain the existence of rigid punishment based partly on outcomes. The model is compelling and it is grounded in a more general, widely accepted view of the role of punishment in the evolution of pro-sociality. If the model is correct, our sensitivity to resultant luck exists because of its unique ability to facilitate deterrence. That is, our sensitivity is based on the very same consequentialist grounds that justify that sensitivity.

Evolutionary theorizing is, to some degree, inherently speculative. However, Cushman's evolutionary model is buttressed by learning studies (Cushman 2013). Remember, according to Cushman, pro-sociality is flexible and is supported by learning mechanisms. Furthermore, the relevant learning mechanisms require causal models of the world—internal models of the causal relationships between actions and outcomes. If this is the way in which people learn to avoid anti-social behavior, Cushman argues, then to punish only their intentions—to ignore their modeling of outcomes—is to miss an important teaching opportunity. People will be more likely to modify their behavior if they are made to understand the contingencies between actions, outcomes, and punishment. So, given these facts about the psychology of learning, we should expect that punishment based partly on outcomes will facilitate learning.

Cushman's experimental research on learning and punishment supports his learning model, by confirming that as learners we are attuned to punishment of outcomes (Cushman and Costa in prep). The research suggests that we are often disposed to modify our behavior, in response to punishment, when punishment is matched to the outcomes of our behavior, more so than when it is matched to the intentions that produce our behavior. In Cushman's study, pairs of participants play a game in which one participant, Shooter, throws darts at a board, while another, Trainer, wins or loses money depending on which targets are hit. Shooter, however, is not told initially which targets will win Trainer money, and which targets will cost her money. The only way for Shooter to learn is through Trainer rewarding and punishing him after each throw, by adding to or subtracting from his total payoff. Before each throw, Shooter announces which target he is aiming for. In half the trials Trainer rewards and punishes Shooter on the basis of what he aims for (intentions), in the other half on the basis of what target he actually hits (outcomes). Cushman and Costa find that participants in Shooter's role are twice as good at learning which targets benefit participants in Trainer's role when the latter's punishments are based on his hits rather than aims—on the outcomes of his actions rather than his intentions.

What does Cushman's work tell us about the philosophy of moral luck? We have, it seems, evolved to become better capable of *moral learning* when punishment is matched to outcomes rather than intentions. Thus, we find a consequentialist rationale embodied in the explanation for our sensitivity to moral luck. Punishment is more likely to deter immoral behavior if it is based partly on outcomes. The explanation of our sensitivity to moral luck mirrors the structure of a two-level, indirect-consequentialist rationale for moral luck. Proximally, we assign blame and punishment on the basis of what is merited by an agent's intentions and the outcomes of her actions. But the ultimate reason that we assign responsibility in this way is that this reliably produced good consequences.

So, it's not the case simply that there are good reasons available for moral luck. Rather, our practices embody moral luck *for* these good reasons. And so, too, the consequentialist rationale for moral luck is not simply a "just-why" story. It's likely that a sensitivity to luck genuinely produces the good consequences that seem to justify it, since the ability of our practices to produce these good consequences explains why they exist (see Kumar 2017 for further discussion).

To be clear, I have not argued that moral luck is real merely on the grounds that our sensitivity to luck is evolved. 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. Rather, I argued (in section five) that moral luck is real because there are strong consequentialist reasons to blame and punish people on the basis not just of their intentions but also the outcome of their actions and inactions. Moral luck is part of an effective method for deterring immoral behavior. Evolutionary theory comes into play by supporting this philosophical argument, that is, by showing that it is not an untrustworthy rationalization of the status quo. We think and act as though morality is subject to luck *because* our practice has good consequences. Of course, it is not inevitable that evolution has good consequences.

Yet, in this case, I have argued, our practices of assigning responsibility partly on the basis of luck did evolve because they have good consequences.

Some philosophers, sympathetic to the evolutionary explanation of moral luck offered here, may find themselves drawn to a different philosophical conclusion. They might plump for an error theory of moral luck. That is, they might agree that there is a consequentialist rationale for moral luck but argue that this contradicts the retributivist nature of our intuitions: we feel as if responsibility is a matter of desert or merit for one's deeds, but we are in error about that since, in fact, its rationale lies in consequences.

A realist view of moral luck is preferable to an error theory because it saves the phenomena. It preserves the intuitions with which this essay began, viz., that some people are genuinely blameworthy for things beyond their control. Whether realism or error theory is, on balance, the most attractive view also turns on other philosophical costs and benefits. For example, can indirect consequentialism can avoid collapsing into direct consequentialism? I have not attempted to answer such broad objections. But I have shown, at least, that an indirect consequentialist theory of blame and punishment is plausible, that it explains why moral responsibility is partly a matter of luck, and that the evolutionary history of blame and punishment protects this theory from the objection that it is unreliable rationalization.

8. Conclusion

This essay has painted a picture of moral responsibility that substantiates resultant moral luck. First of all, skepticism about moral luck conflicts with powerful intuitions about responsibility that underlie our practices of assigning blame and punishment. Furthermore, if we accept that blame and punishment have two rationales—not just retribution but also deterrence—it makes sense to blame and punish on the basis not just of intentions but also outcomes. Outcomes are easier to identify than intentions, and so blame and punishment based partly on outcomes is more reliable, and thus a more effective method of deterrence.

Empirical research plays a role in my case for moral luck. One body of research suggests that our commitment to the control condition is partial and thus suggests a restricted version of the control condition. That is, intentions matters for wrongness, impermissibility, and character and they also matter, but only partially, for blame and punishment. Another body of research challenges the empirical claim made by proponents of the epistemic view that we treat outcomes only as evidence of others' intentions—an attempt to reconcile our practices with the control condition. In fact, we treat outcomes as bearing directly on blame and punishment.

More positively, evolutionary modeling and experimental studies empirically vindicate moral luck. As a whole, this research explains how the practice of assigning responsibility based partly and rigidly on outcomes secured the positive deterrence effects that partly justify our practices. The indirect consequentialist rationale for moral luck is no mere fantasy. It is the reason for our sensitivity to moral luck.

Acknowledgements

I am grateful to many colleagues for helpful discussion of this essay: Mara Bollard, Paul Boswell, Sarah Buss, Richmond Campbell, Fiery Cushman, David Dick, Joshua Greene, Neil Levy, Theresa Lopez, Alice MacLachlan, Michael McKenna, Shaun Nichols, Peter Railton, David Shoemaker, Evan Tiffany, and Jennifer Woodrow. I would also like to thank audiences at the Pacific Division of the American Philosophical Association, the University of Michigan, the Moral Psychology Research Group, the Canadian Philosophical Association, the Society for Philosophy and Psychology, and York University.

References

Adams, R. 1985. “Involuntary sins,” *The Philosophical Review*, 94: 3–31.

Boyd, R. & Richerson, P. 1992. “Punishment allows the evolution of cooperation (or anything else) in sizable groups,” *Ethology and Sociobiology*, 13: 171–95.

Carlsmith, K. 2006. “The roles of retribution and utility in determining punishment,” *Journal of Experimental and Social Psychology*, 42: 437–51.

———. 2008. “On justifying punishment: the discrepancy between words and actions,” *Social Justice Research*, 21: 119–37.

Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge* (Oxford University Press).

Cushman, F. 2008. “Crime and punishment: differential reliance on causal and intentional information for different classes of moral judgment,” *Cognition* 108 (2): 353–80.

———. 2011. “Should the law depend on luck?” in M. Brockman, Ed., *Future Science: 19 Essays from the Cutting Edge* (Vintage).

———. 2013. “The role of learning in punishment, prosociality, and human uniqueness,” in R. Joyce, K. Sterelny, B. Calcott & B. Fraser, Eds., *Signaling, Commitment and Emotion, Vol. 2: Psychological and Environmental Foundations of Cooperation* (MIT Press).

———. 2015. “Punishment: from intuitions to institutions,” *Philosophy Compass*, 10 (2): 117–33.

Cushman, F., Sheketoff, R., Wharton, S. & Carey, S. 2013. “The development of intent-based moral judgment,” *Cognition*, 127: 6–21.

Domsky, D. 2004. “There is no door: Finally solving the problem of moral luck,” *The Journal of Philosophy*, 101: 445–64.

Fehr, E. & Gachter, S. 2002. “Altruistic punishment in humans,” *Nature*, 415: 137–40.

Greene, J. 2008. “The secret joke of Kant’s soul,” in in W. Sinnott-Armstrong, Ed., *Moral Psychology Vol. 3: The Neuroscience of Morality* (MIT Press): 35–79.

Hart, H. L. A. 1968. *Punishment and Responsibility* (Oxford University Press).

Hooker, B. 2016. “Rule Consequentialism,” in E. Zalta, Ed., *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition).

Kumar, V. 2017. “Moral vindications,” *Cognition*, 176: 124–34.

Kumar, V. & Campbell, R. 2018. *Why We Are Moral*. Manuscript in prep.

Lench, H., Domsky, D., Smallman, R., & Darbor, K. 2015. “Beliefs in moral luck: When and why blame hinges on luck,” *British Journal of Psychology*, 106: 272–287.

Levy, N. 2016. “Dissolving the puzzle of resultant moral luck,” *Review of Philosophy and Psychology*, 7: 127–139.

MacKenzie, J. 2017. “Agent-regret and the social practice of moral luck,” *Res Philosophica*, 94: 95–117.

Martin, J. & Cushman, F. 2016. “The adaptive logic of moral luck,” in J. Sytsma & W. Buckwalter, Eds., *The Blackwell Companion to Experimental Philosophy*.

Moore, M. 1997. *Placing Blame: A Theory of Criminal Law* (Oxford: Clarendon Press).

———. 2009. *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics* (Oxford: Oxford University Press).

Nadelhoffer, T., Heshmati, S., Kaplan, D. & Nichols, S. 2013. "Folk retributivism and the communication confound," *Economics and Philosophy*, 29: 235–61.

Nagel, T. 1979. "Moral luck," in *Mortal Questions* (Cambridge: Cambridge University Press).

Nelkin, D. 2013. "Moral luck," *The Stanford Encyclopedia of Philosophy*, E. Zalta (ed.), http://plato.stanford.edu/archives/sum2013/entries/moral-luck/

Rawls, J. 1955. "Two concepts of rules," *The Philosophical Review*, 64: 3–32.

Richards, N. 1986. "Luck and desert," *Mind*, 65: 198–209.

Rosebury, B. 1995. "Moral responsibility and moral luck," *The Philosophical Review*, 104 (4): 499–524.

Royzman, E. & Kumar, R. 2004. "Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck," *Ratio*, 17: 329–44.

Scanlon, T. 2008. *Moral Dimensions: Permissibilty, Meaning, Blame* (Harvard University Press).

Smith, A. 1759. *The Theory of Moral Sentiments*.

Thomson, J. J. 1993. "Morality and bad luck," in D. Statman, Ed., *Moral Luck* (Albany: SUNY Press).

Walker, M. 1991. "Moral luck and the virtues of impure agency," *Metaphilosophy*, 22: 14–27.

Wolf, S. 2001. "The moral of moral luck," *Philosophic Exchange*, 31: 4–19.

Woodrow, J. 2006. *Luck and Responsibility* (PhD Dissertation: Dalhousie University).

Young, L., Nichols, S. & Saxe, R. 2010. "Investigating the neural and cognitive basis of moral luck: it's not what you do but what you know," in J. Knobe, T. Lombrozo & E. Machery, Eds., *Review of Philosophy and Psychology*, Special Issue on Psychology and Experimental Philosophy, 1, 333–49.

Zimmerman, M. 2002. "Taking luck seriously," *Journal of Philosophy*, 99: 553–76.