

Free Will, Self-Creation, and the Paradox of Moral Luck

KRISTIN M. MICKELSON

1. INTRODUCTION

Prior to philosophical inquiry, many people have the intuition that the following moral principle—commonly known as *the control principle*—is unquestionably true: people cannot bear moral responsibility for things which are not their fault, that is, things which are beyond their control (cf. Nagel 1976, 138). We also generally assume that normal people are sometimes morally responsible for their actions, as our everyday moral judgments and practices of blame and praise imply.¹ At first blush, these commonsense positions seem compatible. However, when we carefully consider concrete cases of human action, it becomes plain that every action a person performs and the effects that one's actions have on the world are largely—perhaps entirely—a matter of luck, that is, due to factors which are entirely beyond the actor's control.² As such, it seems that commonsense views about morality and moral responsibility commit us to the existence of something that is, by our own intuitive lights, impossible: the existence of *moral luck*, that is, cases in which a person is morally responsible for something which was not completely under that person's control. The intuitive judgements which generate our apparent commitment to the existence of moral luck seem undeniable, but granting its existence is unacceptable; this is the *paradox of moral luck*.

1. Here, the phrase “moral responsibility” is shorthand for what Derk Pereboom calls “basic desert moral responsibility” (Pereboom 2014, 2).

2. I will use the phrase “due to luck” as a shorthand for “due to factors beyond one's control,” for it seems that the “lack of control” account of luck captures the notion of luck which is central to the moral-luck debate (cf. Anderson 2019, Hartman 2017, 23–31, Statman 2019).

In this essay, I argue that moral luck is genuinely paradoxical, and propose a solution to the paradox of moral luck which implies that moral luck does not exist. To make my case, I highlight a few key theoretical connections between the paradox of moral luck and two related problems, namely *the problem of free will and determinism* and *the paradox of self-creation*. This essay has four main sections. I begin (§2) with a discussion of the traditional problem of free will and determinism. In the light of recent developments in the free-will literature, I argue that the free-will/determinism debate covers all three of the basic categories of luck discussed in the moral-luck literature, namely: (i) *causal-relation luck*, which obtains when one's action is due, at least in part, to causal relations that are beyond one's control;³ (ii) *noncausal/actor-extrinsic circumstantial luck*, which obtains when one's action is due, at least in part, to circumstances (events, states of affairs, etc.) external to the actor that are beyond that actor's control; and (iii) *constitutive luck*, which obtains when one's actions are due, at least in part, to constitutive properties of the actor that are beyond the actor's control.⁴ Indeed, I argue that there is good reason to think that the upshot of the free-will/determinism debate is that free will is impossible even though causal relations (laws of nature, etc.) play *no role whatsoever* in robbing people of free will when determinism is true. If this is right, we must reject the standard narrative in the moral-luck literature

3. Here, I break from the normal characterization of causal luck as the *conjunction* of luck in the causal relations which hold (e.g., whether they are deterministic or probabilistic) and luck in antecedent states of affairs. That is, I assume that causal relations seem to belong to a distinct ontological category of "factors beyond our control which play a key role in settling what happens to us." In doing so, I grant a familiar background assumption of the contemporary free-will debate, namely that states of affairs and the causal relations between them are distinct phenomena which can come apart (e.g., that two worlds may be exact duplicates at one timeslice but not others in virtue of the fact that different causal relations—causal laws, laws of nature, and so on—obtain in the two worlds). I also promote a taxonomy which allows us to debate (and perhaps crisply deny) that our lack of control over causal relations *per se* affect one's status as a free and morally responsible agent. Notably, this conflicts with Latus's suggestion that the entire category of causal luck is redundant on the grounds that "Constitutive luck covers what we are, while circumstantial luck covers what happens to us" (Latus 2001, echoing Nagel 1986, 113–14; Levy [2011, 40] makes a similar claim, but his argument seems to turn on the definition of "luck").

4. The three ontological categories of luck listed here are implied by Thomas Nagel (1976), but these are not the categories he explicitly identifies. Nagel characterizes the free-will debate as a debate about antecedent "causes of action," and contrasts this with the classes of "constitutive luck," "luck in one's circumstances—the kind of problems and situations one faces," and luck in the "effects of action" (Nagel 1976, 140). However, what qualifies as a "causal factor" or "circumstantial factor" is not stable in Nagel's work. For example, both constitutive properties of the actor and noncausal/actor-extrinsic circumstances are often treated as causes of action (i.e., as "casual factors," broadly construed; cf. Nagel 1976, 1979, 1986); other times, constitutive luck, causal relations, and noncausal/actor-extrinsic circumstances seem to qualify as "circumstantial factors" (cf. Nagel 1986, 174); and other times circumstantial luck is very narrowly defined as "the kind of problems and situations one faces" (Nagel 1976, 140), suggesting that the category is limited to something like "tricky situations" (cf. Herdova and Kearns 2017). Moreover, Nagel's "fourth category," that is, luck in effects of an action, is based on when the luck occurs relative to a given action and, so, is not a contrast class to the three forms of luck I have listed. Overall, Nagel's "four categories" of luck seem more like four rhetorical "centers of gravity," that is, rhetorically interesting combinations of the three factors beyond our control which I have listed, where different cases can be developed based on when the specified factors obtain relative to a given action.

according to which the problem of free will and determinism is a narrow debate about whether causal-relation luck—perhaps in conjunction with circumstantial luck⁵—precludes free will (cf. Hartman 2017, Latus 2001, Nelkin 2019, Sartorio 2019).⁶ It also implies that we need a noncausal explanation for what makes it impossible—at least for beings like us—to act freely when determinism is true.

In pursuit of such an explanation, I offer (§3) an overview of *manipulation arguments*. These arguments are typically classified as “arguments for incompatibilism,” but I explain how one might develop a manipulation argument in support of the anti-incompatibilist conclusion that free will is impossible—and, so, impossible when determinism is true—even though deterministic causal relations pose no threat whatsoever to free will. With that negative conclusion in hand, I turn (§4) to a problem that arises in both the free-will literature and the moral-luck literature, but is generally sidelined in both: the paradox of self-creation.⁷ By clarifying the unique dialectic set up by the paradox of self-creation and drawing out its intimate connections to the problem of free will and determinism, I show that the paradox of self-creation is arguably at the heart of the problem of free will and determinism. That is, I propose that the debate over the relationship between free will and determinism is, at its core, a debate about whether constitutive luck makes people unfree.

Finally, I draw (§5) on my characterizations of the paradox of self-creation and the problem of free will and determinism to shed new light on the paradox of moral luck. I argue that these three problems are just different rhetorical frameworks through which to investigate the same basic metaphysical problem, namely that of positively identifying the type of control one must have in order

5. The apt phrase “remote causal luck” is already in use for luck resulting from the conjunction of circumstances in the remote past and causal-relations; the apt phrase “remote deterministic luck” (Mele 2006, 77) is already in use for luck resulting from the conjunction of circumstances in the remote past and deterministic causal relations.

6. Nagel’s suggestion (1976, 140, 148) that the free-will/determinism debate is focused on “causes of action” (along with the use of phrases such as “determinism obliterates responsibility” [Nagel 1976, 150]) has contributed to the idea that the free-will debate is either narrowly about causal relations per se, or the conjunction of causal relations and certain states of affairs (a.k.a. remote deterministic luck). As I argue in the final section of this paper (§5), Nagel’s writings do not commit him to such a narrow view of the free-will debate and, indeed, much of what Nagel says fits well with my contention that the free-will/determinism debate covers all three forms of antecedent luck that I have identified.

7. The problem of self-creation is discussed more often and more rigorously in mainstream moral-luck literature than in mainstream free-will literature. Still, philosophers in the moral-luck debate often suggest that the unavoidability or even incoherence of essential constitutive luck provides reason to think that it is not a serious threat to the control required for moral responsibility (cf. Greco 1995, 95, and reply by Latus 2000, 169; Rescher 1995, and replies by Latus 2003 and Nelkin 2019). Such replies seem dialectically infelicitous in discussions of the paradox of moral luck, for apparent cases of moral luck give us reason to *expect* that we will find something incoherent or otherwise fundamentally problematic in our ordinary notion of moral responsibility. Put another way, apparent moral luck cases raise the worry that, when it comes to our intuitive commitments regarding moral responsibility, what we want is impossible (Nagel 1986, 117; Arpaly 2006, Ch. 4). So, the fact that it is impossible to have control over our essential properties speaks for rather than against it as a diagnosis of what drives our paradoxical intuitions regarding moral luck.

to satisfy the control condition on moral responsibility.⁸ Although the categories of causal-relation luck, circumstantial luck, and constitutive luck are not unique to the moral-luck debate,⁹ the temporal distinction between whether the luck occurs before or after one's action is. Drawing the temporal distinction between *antecedent luck* and *posterior luck* (cf. Williams and Nagel 1976)¹⁰ helps us to harness the unique rhetorical force of apparent cases of posterior moral luck—for example, standard “truck driver” cases in which whether a driver is morally responsible for killing a child or not depends on factors beyond the driver's control, namely whether/when a child runs into the street in front of them.¹¹ Cases of apparent posterior moral luck are uniquely compelling ways to raise the worry that there is a serious problem buried in our ordinary conception of moral responsibility and to motivate the need to identify that problem. Ultimately, though, it seems the same fundamental problem is highlighted by apparent anterior and posterior moral luck cases, determinism scenarios, and the paradox of self-creation.¹² I precisely identify this problem in a *source-paradox solution* to the problem

8. This suggestion has been explicitly raised and rejected by at least one influential figure in the moral-luck debate (cf. Latus 2001).

9. All three are covered in the free-will literature; constitutive luck is the central topic of the self-creation paradox literature.

10. Posterior luck is generally portrayed as a *fourth* category of luck (see footnote 4), as if it were a contrast class to constitutive luck, causal-relation luck, and (noncausal/actor-extrinsic) circumstantial luck. (The terms “resultant luck” [Zimmerman 1987, 376] and “consequential luck” [cf. Enoch and Marmor 2006] are generally tied narrowly to posterior circumstantial luck.) However, it seems that there is space for both posterior causal and posterior constitutive luck. Posterior constitutive luck seems to be at play in discussions which focus on our lack of control over the constitutive properties we acquire as the result of performing a given action, as in the cases of “transformative experiences” (cf. Herdova 2019) and discussions of Oedipus's “moral taint” (cf. Hartman 2017, 136–27). Posterior causal-relation luck also seems coherent, even if it is unclear whether there is any way to develop a compelling case of (apparent) posterior causal *moral* luck. In short, it seems reasonable to think that the three basic ontological varieties as luck—constitutive, causal-relation, or circumstantial—can be present both before and after any given action. Neal Tognazzini (2011, 98, n. 6) makes a similar point, but his proposed taxonomy does not isolate causal-relation luck and also misleadingly implies that constitutive and circumstantial luck are contrast classes to antecedent luck and resultant luck. Perhaps one could plausibly argue that posterior luck is a fourth category of luck by claiming that it (unlike antecedent luck) results from our epistemic limitations in a way that antecedent luck does not (cf. Levy 2016, Hartman 2017, Herdova 2019), but I will not address this position here.

11. The original “lorry driver” case of posterior luck is from Bernard Williams (1976). For a review of refined versions of such cases and their import, see Hartman (2017).

12. Notably, Williams's original work on moral luck focused primarily on posterior luck and posterior moral-luck cases, and Nagel criticized Williams for this (Nagel 1976, 140). However, it is not clear that Nagel's criticism hits its mark, for only apparent cases of posterior moral luck are unique to the moral-luck literature and they seem to be the most intuitively compelling (cf. Hartman 2017, 129). Posterior moral luck cases are especially important because they get traction with philosophers who, given their subjective interests, are more drawn to tensions in familiar moral principles (e.g., the control principle) and ordinary post-action practices of blame and praise than to seemingly esoteric metaphysical inquiries into the nature of blameworthiness and praiseworthiness itself. However, we can grant that Williams was right to focus on posterior luck in establishing the paradox of moral luck as a distinct area for future research, and also accept—as I will argue we should—that Nagel is right when he says that the different kinds of moral luck “present a common problem” (Nagel 1976, 140).

of moral luck. In brief, this solution states that moral responsibility requires a certain type of basic control over one's actions, and this type of basic control, in turn, can only be exercised by someone who is a genuine originator or source of his own actions. We generally take ourselves and others to satisfy this source condition, not knowing quite what it amounts to. Upon review, however, there is no coherent way to spell out the requisite type of "sourcehood"; the sourcehood we want—and generally take ourselves to have—is metaphysically impossible. This source-paradox solution explains why the problem of moral luck arises, while also explaining why moral luck does not exist.

Although I defend a skeptical solution to the problem of moral luck in this essay, my central goal is not to persuade the reader to embrace my source-paradox solution or, more generally, to accept that moral luck does not exist. My main aim, rather, is to highlight a few overlooked connections between the problem of free will and determinism, the problem of self-creation, and the problem of moral luck, and to suggest that more explicit and rigorous work on these connections is warranted.

2. THE PROBLEM OF DETERMINISM SCENARIOS

The traditional problem of free will and determinism is often characterized as a dispute about the relationship between deterministic causation (deterministic laws of nature, deterministic evolution, or the like) and free will, where "free will" picks out the type of control that one must exercise in order to be morally responsible for one's actions.¹³ Contemporary philosophers working on the paradox of moral luck generally assume that the problem of free will and determinism is a narrow dispute about whether deterministic causal relations—perhaps in conjunction with states of affairs prior to the action—preclude free will and, thereby, moral responsibility.¹⁴ After pigeon-holing the problem of free will and determinism in this way,

13. Because the free-will debate is primarily a debate about the nature of free will itself, it is difficult for philosophers to provide a precise yet reasonably uncontroversial characterization of free will: the more precise one's proposed characterization of free will, the more likely that one camp or another will find it question-begging; the more generic the proposal, the more likely that the referent of 'free will' will not be fixed securely enough to ensure that the interlocutors in the debate are disagreeing about the same thing (rather than having a mere verbal dispute). It is now common to fix the referent of "free will" by tethering the term to the control condition on moral responsibility. For purposes of this paper, I will fix the referent of "free will" by saying that one has free will if and only if one satisfies the necessary control condition on moral responsibility. I will use the phrase "free will" interchangeably with "the control condition on moral responsibility," but I am not committed to the view that they are indistinguishable.

14. I use the term "precludes" to indicate an asymmetrical antagonistic relevance relation, such that "A precludes B" only if A makes it the case that B does not obtain. (Notably, this means that "A precludes B" and "B precludes A" are fundamentally different claims; I take no stand here whether there are cases in which A precludes B and B also precludes A.) This is worth noting because several philosophers (in private conversation/correspondence) have informed me that they sometimes use "preclude" to pick out a symmetrical, nonrelevance relation. For example, some philosophers define "precludes" so that "A precludes B" is true when A is spuriously impossible with B.

philosophers typically sideline the problem from their discussion of moral luck.¹⁵ In this section, I argue that it is wrong to characterize the problem of free will and determinism as a narrow debate about whether deterministic causal relations (deterministic laws, deterministic evolution, etc.) undermine free will.

Many debates now fall under the umbrella of “the problem of free will and determinism,” and the narrow question of whether causal relations pose a threat to free will is undeniably one of them. However, for purposes of clarifying the connection between the problems of free will and moral luck, we must locate this question in the broader investigation of free will vis-à-vis determinism. I will do this by focusing on *the problem of free will and determinism scenarios*, roughly a debate in which the hypothesis of determinism is used to motivate an open-ended investigation into the nature of free will itself.

The problem of free will and determinism scenarios begins by considering a world of a special sort. In this world, there are features—of the sort commonly associated with causation, laws of nature, and so on.—which account for the state-by-state evolution of the physical world. More specifically, these features are such that they, together with noncausal circumstances (events, states of affairs, etc.) in the remote past, unconditionally fix one unique future; the future is just an “unfolding of the arbitrarily given” (Smilansky 2003, 275).¹⁶ A world which fits this definite description is a world at which determinism is true.¹⁷

Determinism scenarios are jarring in part because they force us out of our normal first-person experience of the world, allowing us see how we—and all human actions—are embedded in the grand causal nexus of the natural

15. For example, Robert Hartman (2017, 4–5) defends the practice of sidelining the problem of free will and determinism in discussions of moral luck, even though he later provides several laudably rigorous discussions of the connections between compatibilist and libertarian-friendly accounts of free will and moral luck. Other philosophers have worked on bridging the divide between the moral-luck literature and the free-will literature as well (cf. Tognazzini 2011), but (to my knowledge) the present essay is the first full-fledged attempt to isolate the distinct dialectical frameworks around the problem of moral luck, the problem of free will and determinism scenarios, and the problem of self-creation while also highlighting the need to provide a unified response to all three.

16. The term “remote past” is often a technical term in the free-will debate, referring roughly to those times at which the universe existed prior to the existence of the first human beings (cf. Campbell 2007).

17. My use of “determinism” corresponds to its traditional use in the free-will debate, following the metaphysical characterization of the doctrine provided by William James (1884). While there may be some dialectical contexts in which the definition of “determinism” is up for substantive debate, the problem of determinism scenarios is based on a thought experiment which consists partly in carefully stipulated details about the evolution of the world; the name we use to refer to evolution of the relevant kind—and such things as whether or not the description of the evolution of the world provided in the thought experiment is “scientifically respectable” or consistent with the true theory of the causation/laws of nature—is irrelevant in the narrow dialectical context surrounding the correlation and explanation problems of free will and determinism scenarios.

world (cf. Nagel 1976, 1979, 1986). From this external standpoint, we can see factors which were not up to us—such as the laws of nature and facts in the remote past—make each of our actions inevitable. As William James poetically expresses the worry: “The whole sting and excitement of our voluntary life ... depends on our sense that in it things are really being decided from one moment to another, and that it is not the dull rattling off of a chain that was forged innumerable ages ago” (James 1890, 453). It is unsurprising that determinism scenarios commonly elicit the intuition that people in such scenarios can have no more control over their actions than they have over the laws and the remote past, which is to say that they can have no control at all.

A token human acting in a determinism scenario is just an arbitrary representative of all normal humans, so the intuitive judgement that the normal human in a determinism scenario lacks free will and moral responsibility supports the more general and bolder claim that it is *metaphysically impossible* for any normal human to act freely in a normal determinism scenario. For many philosophers, this intuitive judgement—hereafter the *victim intuition*—is so strong and clear that it provides a data point which any viable theory of free will and, so, moral responsibility, must accommodate. For philosophers who have this type of “data-point” victim intuition, determinism scenarios seem to expose a desideratum on any viable theory of free will: it must include at least one necessary condition which cannot be satisfied (by a normal human at least) in a normal determinism scenario. One reason that determinism scenarios generate a serious philosophical problem is that many people who reflect upon such scenarios have a data-point victim intuition which constrains their future theorizing about free will, but many people do not. Those in the latter camp may deny that a viable account of free will must pinpoint a necessary condition on free will which cannot be satisfied in a normal determinism scenario. As such, whether one has or lacks a data-point victim intuition to determinism scenarios marks a fundamental divide in the free-will debate—a divide that reflects not only a disagreement about what free will is, but reveals that philosophers fundamentally disagree about what must be achieved by an adequate theory of free will.¹⁸

In order to understand the full dialectic surrounding the problem of free will and determinism scenarios, it is essential to recognize that a victim intuition is merely a judgment about the status of a human actor, namely whether the normal human in the determinism scenario *has* or *lacks* free will. A victim intuition, taken on its own, does not reveal precisely *why* humans lack free will in normal determinism scenarios. Once we appreciate this, we can tease apart two commonly conflated problems, what I call the *correlation problem of free will and determinism scenarios* and the *explanation problem of free will and determinism scenarios*.

18. Elsewhere I call this intuition-driven dispute over the desideratum on free will the *common-ground crisis* (Mickelson forthcoming-b), for there seems to be an important dialectical stalemate around the (in)compossibilism divide.

The correlation problem of free will and determinism scenarios may be summarized as the question: “Is it metaphysically possible for a normal human to exercise free will, that is, moral-responsibility-relevant control, in a normal determinism scenario?” Those with a victim intuition answer “no,” thereby committing themselves to the view that deterministic evolution (laws, causation, etc.) and free human agents are impossible, that is, it is impossible for these phenomena to co-exist.¹⁹ The latter view, hereafter *incompossibilism*, is one candidate solution to the correlation problem; if *incompossibilism* is false, then *compossibilism* is true.

Notably, though, just as two things may be correlated without bearing a direct causal relation to each other, phenomena may be impossible even when there is no deeper relevance relation between them. That is, two things may be metaphysically impossible even though they are not metaphysically incompatible. The latter relation holds only when there is an antagonistic explanatory/relevance relation grounded in the respective natures of the impossible things and this is what makes it the case that the two things cannot co-exist.²⁰ For example, round squares are impossible with fluffy kittens, since their impossibility follows *a fortiori* from the impossibility of round squares; there is no conflict in the nature of fluffy kittens and the nature of round squares such that the existence of fluffy kittens keeps round squares from existing (or vice versa). Borrowing from the empirical sciences, let us say that an impossibility relation is *spurious* when it holds in the absence of a metaphysical incompatibility relation. The key point here is that a philosopher may accept that people *are* unfree in determinism scenarios without embracing a particular view about what *makes* people unfree in the determinism scenarios.

Once a philosopher embraces *incompossibilism*, they face a distinct challenge of accounting for the lack of free human agency in determinism scenarios. To answer this new challenge, the *incompossibilist* must solve a two-part *explanation problem of free will and determinism scenarios*:

The Unmet Condition Problem (E1): The *incompossibilist* must pinpoint the necessary condition C which goes unmet in normal determinism scenarios;

The Underminer Problem (E2): The *incompossibilist* must pinpoint what keeps people from satisfying C in normal determinism scenarios.

19. More carefully, this true so long as we assume that we are talking about normal humans, where “normal” humans have a remote personal past (cf. Campbell 2007).

20. Likewise, propositions may be (necessarily) logically inconsistent without being (necessarily) logically incompatible. Notably, Keith Lehrer, who seems to have coined the term “*incompatibilism*” in his dissertation, introduced the term to pick out the view that the thesis of determinism is *logically incompatible* with the thesis that a normal human has free will. Lehrer has confirmed (in correspondence 2019) that the notion of logical incompatibility he had in mind when he coined the term was not mere logical inconsistency, that is, “not just material conditional truth, but the truth of the material conditional resulting from logical and semantic relations alone.”

(E2) has not been widely recognized as a distinct challenge. Rather, it has been generally taken for granted that whatever an impossibilist's preferred solution to (E1), the impossibilist must accept the same solution to (E2). Specifically, philosophers generally assume that impossibilists must accept that deterministic causal relations (deterministic laws of nature, deterministic evolution, or the like) are metaphysically incompatible with free will.²¹ Hereafter, I will use the label "incompatibilism" to pick out the narrow explanatory view that the type of global deterministic causation described by determinism is metaphysically incompatible with—precludes, undermines, destroys, makes impossible—free will; "compatibilism" will name the view that both incompatibilism and its corollary, impossibilism, are false.²²

Because philosophers have generally failed to recognize that impossibilism may be true even when incompatibilism is false, the traditional in-house debate among impossibilists has focused almost exclusively on their rival solutions to (E1). Given the aims of this paper, one trend in proposed answers to (E1) is worthy of attention. One traditional answer to (E1)—promoted by eminent philosophers including Peter van Inwagen (1983)—is that free will has a certain type of "could-have-done-otherwise" condition which cannot be satisfied when determinism is true.²³ According to these impossibilists, the could-have-done-otherwise condition must be understood as an *actual-sequence* ability to do otherwise. Roughly, the idea is that the future is like a garden of forking paths and when we choose one of these paths over the others, this choice was free only if we could have done otherwise had the same circumstances leading up to that choice been exactly the same (cf. van Inwagen 1990). However, in the wake of challenges to the could-have-done-otherwise solution to (E1),²⁴ some prominent impossibilists have argued that the relation between free will and the ability to do otherwise is more complicated than traditionally thought. For example, Robert Kane argues that free will has a complicated *source condition* and one must exercise an actual-sequence ability to do otherwise to satisfy this sourcehood condition. Kane, in other

21. Kadri Vihvelin (2008, 2013) and Neil Levy (2011) are among the few philosophers who have drawn attention to the distinction between (what I call) impossibility and metaphysical incompatibility, but they do not identify the distinct dialectical value of the modest impossibility claim on its own (cf. Mickelson 2015a).

22. For clarity, let me summarize the key relations between views in the proposed taxonomy. Impossibilism does not entail incompatibilism; impossibilism is a corollary (rather than a defining tenet) of incompatibilism. Compatibilism is the conjunction of compossibilism (i.e., the negation of impossibilism) and *anti-incompatibilism*, the negation of the incompatibilism. I will gloss over the compossibilism/compatibilism distinction here since it is not salient to the thesis of this paper. For further discussion of the motivation for this taxonomy, see Mickelson (2015a, 2015b, forthcoming-a, forthcoming-b).

23. Compatibilists may grant that there is a could-have-done-otherwise condition on free will (cf. the discussion of "classical compatibilism" in McKenna and Pereboom 2016), but they must flesh out this condition so that it can be satisfied by someone living in a normal determinism scenario.

24. Arguably the biggest motivation for the shift was Harry Frankfurt's development of so-called "Frankfurt-style cases" (Frankfurt 1969).

words, provides an answer to (E1) in which an actual-sequence could-have-done-otherwise requirement is nested in the source condition on free will (cf. Kane 1996, 2002, 2004).²⁵

In contrast to the above solutions to (E1), other prominent impossibilists have adopted a pure source-condition solution to (E1). These pure source impossibilists argue that free will has a necessary source condition, but they explicitly reject the proposal that it has an actual-sequence could-have-done-otherwise condition (either as a free-standing necessary condition or embedded in the source condition). For instance, Derk Pereboom (2001, 2014) and Galen Strawson (1986, 1989, 1994, 2008) are both source-only impossibilists in this sense. Because Pereboom and Strawson each hold that it is impossible for normal humans to act freely, they are generally lumped together as incompatibilists and free-will skeptics. However, this classification glosses over the fact that these philosophers flesh out the source condition in very different ways, that is, they advance very different source-condition solutions to (E1).²⁶

Pereboom holds that deterministic causation (in conjunction with non-causal circumstances beyond one's control) is metaphysically incompatible with—undermines, precludes, destroys, makes impossible—human free will. Put another way, Pereboom holds that there is a direct antagonistic relationship between free will and deterministic causation, such that the presence of deterministic causation (partly in virtue of being deterministic) is relevant to the lack of free human agents when determinism is true. Pereboom, in short, accepts an incompatibilist solution to (E2). Strawson, however, is not an incompatibilist in this sense.²⁷ Strawson agrees with Pereboom that we lack free will because our actions are ultimately due to factors beyond our control, but Strawson disagrees with Pereboom about which factors beyond our control actually do the “work” of destroying free will. According to Strawson, our lack of control over our own constitutive properties—that is, our constitutive luck—keeps us from satisfying the source condition on free will. To Strawson,

25. Peter van Inwagen also suggests that an actual-sequence ability to do otherwise delivers an unsatisfying answer to (E1), for when someone executes a causally undetermined action it is “not really his act at all” (cf. 1983, 130)—and this sounds very close to the claim that the actual-sequence could-have-done-otherwise condition is a failed attempt to flesh out the type of sourcehood required for free will and moral responsibility. In short, I think it is plausible to think that most, if not all, libertarian-friendly attempts to spell out the actual-sequence could-have-done-otherwise condition can be understood as (failed) attempts to flesh out the source condition on free will.

26. It also glosses over the fact that Pereboom is merely an *anthropocentric* impossibilist (i.e., he holds that it is impossible for a normal human); Pereboom rejects Strawson's unqualified impossibilist view that free will is impossible for any possible being in any possible scenario. This is not merely a taxonomical point. Pereboom's case for incompatibilism seems to be in tension with his rejection of unqualified impossibilism, from which it arguably follows that Pereboom's well-known view “hard (anthropocentric source) incompatibilism” is an untenable view (cf. Mickelson, “Hard Times for Hard Incompatibilism,” ms.).

27. Pereboom characterizes the difference between his incompatibilist view and Strawson's anti-incompatibilist view in roughly the same way (cf. Pereboom 2001, 129).

then, it is constitutive luck—and not at all causal-relation or circumstantial luck—which undermines free will and, by extension, moral responsibility. This means that Strawson is both an impossibilist, insofar as he embraces an impossibilist solution to the correlation problem, and an *anti-incompatibilist*, insofar as he rejects the traditional incompatibilist explanation for the truth of impossibilism (i.e., the incompatibilist solution to [E2]). In short, there are important similarities between Pereboom’s and Strawson’s replies to the explanation problem, but there are also important differences.

We will return to Strawson’s argument for the anti-incompatibilist solution to the explanation problem below. The main lesson I want to draw from our discussion so far is that philosophers who accept an impossibilist solution to the correlation problem may disagree substantively about the solution to the explanation problem. Through the lens of the correlation/explanation problem framework, we can see that impossibilists must use reflective equilibrium to develop a precise solution to (E1) that coheres with their response to (E2). This project leaves ample dialectical space for the development of anti-incompatibilist solutions to (E2). In their effort to develop a cohesive response to (E2), impossibilists may appeal to any combination of the three basic ontological categories of antecedent luck recognized in the moral-luck literature, namely causal-relation luck, (noncausal/agent-extrinsic) circumstantial luck, and constitutive luck. In other words, the problem of determinism scenarios is roughly equivalent to the general problem of antecedent luck, rather than the more narrow problem of antecedent causal-relation luck. Against this backdrop, it seems the standard practice of equating the problem of free will and determinism with the problem of deterministic causal-relation luck should be abandoned, for it conflates the problem raised by determinism scenarios with one of its candidate solutions.²⁸

3. MANIPULATION ARGUMENTS

Our review of the basic logical structure of the problem of determinism has helped us to see that we may grant that there is a threat to free will in determinism scenarios without granting that this threat is (even in part) the deterministic causal relations which obtain in such scenarios. Let us now turn to the bolder claim that the upshot of determinism scenarios may be, among other things, that it is impossible to act freely when determinism is true even

28. It is also commonly assumed that the problem of free will and determinism is distinct from the problem of moral luck because the former is limited to a discussion of whether causal-relation luck completely undermines free will, while the latter is focused narrowly on the question of whether factors beyond one’s control may diminish *without completely undermining* one’s moral responsibility. However, this purported distinction between the problems has corroded over years, as evidenced by the recent uptick in discussions of whether certain types of causal-relation luck might merely “mitigate” free will (cf. Todd 2011, Tierney 2013, and Khoury 2014 on mitigation by deterministic causal relations; for the view that probabilistic causal relations “diminish” but do not undermine free will, see Kane 2004, 85).

though deterministic causal relations pose no threat whatsoever to free will or moral responsibility. Towards that end, let us look at a set of arguments known collectively as “manipulation arguments,” and consider how we might develop this type of argument to show that the incompatibilist explanation for impossibilism is wrong.²⁹

All manipulation arguments have roughly the same underlying logical structure—which includes a counterexample step, a generalization step, and an explanation step—but they may be developed in different ways and used to defend a variety of different conclusions (Mickelson 2017).³⁰ The counterexample step of every manipulation argument consists in a thought experiment involving a manipulation scenario. The details of individual cases vary, but the basic storyline is always roughly the same: a normal human is covertly manipulated by another agent or collection of agents to perform a particular action A. For example, the argument might start off with a manipulation scenario in which neuroscientists have technology which allows them to covertly invoke any mental states they like in their victim, call him Vik. One day, the neuroscientists use their technology to cause Vik to undergo a process of reasoning which ends with him intentionally killing his best friend. Intuitively (or so the argument goes), it seems that Vik could not be morally responsible for killing his friend, for his decision and act of killing his friend was not adequately under Vik’s control—after all, Vik “decided” to kill his friend due to a process of reasoning which was directly caused by the meddling neuroscientists. The proponent of the argument then claims that the intuitive judgement that Vik is not morally responsible because he lacks free will spells trouble for compatibilists, for it seems that Vik can satisfy all of the purported necessary conditions on free will that have been proposed by compatibilists to date.³¹ The conclusion of the counterexample step, then, is that philosophers have not yet delivered an adequate compatibilism-friendly account of free will.

While there are various compatibilist-friendly replies to the counterexample step (cf. Pereboom 2014, ch. 4; Mickelson 2017), let us assume, for the sake of argument, that the counterexample step of our manipulation

29. The earliest of these multiple-case manipulation arguments are Richard Taylor’s “puppet argument” against compatibilism (Taylor 1966) and van Inwagen’s “freakish demon” argument against libertarian-friendly views of free will (1983, 130–34); they were brought into the mainstream by Derk Pereboom’s “four-case argument” (Pereboom 2001, 89). The first explicitly anti-incompatibilist manipulation argument is forwarded in Mickelson (2015b) and is developed in full in my “Mastering the Manipulation Argument,” ms.

30. Manipulation arguments need not flesh out all three steps in the same way—indeed, some do not flesh out the explanation step at all (cf. Mickelson 2015b and 2017). However, a manipulation argument must flesh out the generalization step to some degree (however modest) to qualify as a manipulation argument; without the generalization step, the manipulation case is merely a proposed counterexample to extant compatibilism-friendly accounts of free will.

31. For a useful summary of the targeted compatibilism-friendly conditions, see Pereboom (2014, Ch. 4). Of course, whether the counterexample step succeeds depends on getting the details of the manipulation “just right,” and this is no easy task (cf. Demetriou 2010).

argument has succeeded.³² In the generalization step of standard manipulation arguments, the intuition elicited by the manipulation scenario in the counterexample step is used as the basis for a generalization argument for impossibilism. In most manipulation arguments, this generalization argument consists in a slippery-slope argument in which the details of the manipulation case are changed in seemingly irrelevant ways to construct a new case, that new case is then changed in seemingly irrelevant ways to construct a third case, and so on until the series ends with a normal human living in a normal determinism scenario.³³ These cases lend support to the key claim that there is no principled difference between the victim in the manipulation scenario and a normal human in a determinism scenario. If this “no-difference” claim is true, it supports the following conditional: If the manipulation victim is not free (and, so, not morally responsible), then it is impossible for a normal human to act freely in a normal determinism scenario. If the counterexample step of the manipulation argument is successful, then the antecedent of this conditional is true. So, when the counterexample step and the generalization argument are taken together, they provide an intuitively compelling argument for impossibilism.

Impossibilism is a rhetorically interesting stopping place for a manipulation argument. Among other things, it places a burden of proof squarely on the compatibilist: to save compatibilism, the compatibilist must identify a principled difference between the manipulation scenario and the determinism scenario which stops the generalization. Until an adequate principled difference is positively identified, the impossibilists have the dialectical upper hand.³⁴ That said, it is important not to lose sight of the fact that impossibilism expresses a potentially spurious impossibility claim. Nothing in the manipulation argument presented so far indicates that there is a freedom-relevant difference between the manipulation scenario, determinism scenarios, and indeterminism scenarios. Nor does it suggest that there is a freedom-relevant difference between human and nonhuman actors (e.g., it may or may not be that having human DNA plays a role in making humans unfree when determinism is true). So, just as we could not assume that the manipulation per se was freedom-undermining in the initial manipulation scenario, we cannot simply assume that

32. Standard issues regarding the success conditions for a proposed counterexample arise at this stage (e.g., whether the targeted compatibilist-friendly conditions are in fact met and, if so, whether the target audience may reasonably deny having a victim intuition). For a careful discussion of such replies, see Pereboom (2014, ch. 4).

33. I say “most” for two reasons. First, sometimes no positive argument is given in favor of the no-difference claim on the grounds that merely asserting the no-difference claim adequately shifts the burden to the argument’s target audience (cf. Mele 2006, 189). Second, there are indeterministic versions of such slippery-slope manipulation arguments (cf. van Inwagen 1983, 130–34).

34. That is, until the compatibilist pinpoints the difference, she must substantively lower her credence in the truth of compatibilism.

deterministic causal relations play a role in making normal humans unfree in normal determinism scenarios.

Since there is nothing in the underlying logical structure of manipulation arguments that forces the generalization to stop at impossibilism, stopping the generalization argument at impossibilism is premature unless we have positively identified a reason to think that the generalization argument cannot be extended to all possible scenarios.³⁵ Succinctly put, the standard generalization argument for impossibilism is just an arbitrarily restricted version of the following “master” generalization argument:

1. Victim Premise: Vik lacks free will, that is, the control required for moral responsibility.
2. No-difference Premise: There is no relevant difference between Vik in the manipulation scenario and any other metaphysically possible being in any metaphysically possible scenario—so, if Vik lacks free will, then free will (i.e., the control required for moral responsibility) is metaphysically impossible.
3. Impossibilism: Free will (i.e., the control required for moral responsibility) is metaphysically impossible.

This argument is clearly valid. Assuming, as we have, that the counterexample step was successful, the Victim Premise of this argument is true. As such, the success of the master generalization argument hangs on the No-difference Premise.

The No-difference Premise of the master generalization argument extends the dialectical challenge posed by the standard generalization premise to all possibilists. That is, the No-difference Premise of the master generalization argument generates an *unstoppable generalization challenge*: the possibilist must identify a principled difference between Vik in the manipulation scenario and a possible being in some possible scenario in order to block the generalization.³⁶ Until a principled difference is identified which explains why

35. Pereboom delivers a special type of generalization-with-explanation argument in his famous four-case argument and this generalization argument has the grounds for stopping at impossibilism built in. However, this is not usually how the argument is formally summarized (cf. McKenna 2004)—perhaps out of charity, for the explanation component of Pereboom’s generalization argument may be considered a weakness rather than a strength (cf. Mickelson “Hard Times for Hard Incompatibilism,” ms.).

36. The possibilist may reject the Victim Premise, but to reject this premise is simply to say that the manipulation argument already failed at its counterexample step, and it follows that the generalization argument built upon the initial counterexample step also fails. Assuming (as we have) that the counterexample step was successful and we are at the generalization step, the possibilist must reply directly to the No-difference Premise (given that the validity of the argument is uncontentious).

some possible agents have free will even though manipulated Vik does not, the impossibilists have the dialectical upper hand.³⁷

A slippery-slope generalization argument may do a fair job of ruling out candidate explanations for manipulated Vik's lack of free will, but there is a better way of defending the No-difference Premise if one's ultimate goal is to complete all three steps of the manipulation argument. The final explanation step of a "master" manipulation argument (i.e., any manipulation argument which forwards a master generalization argument in its generalization step) must forward an adequate explanation for the impossibility of free will. In other words, the argument must propose an answer to both (E1) and (E2) of the explanation problem. To set the stage for this explanation step, we can support the No-difference Premise of the generalization argument using a familiar type of abductive argument, one in which we identify distinct independent variables and dependent variables in an effort to tease out precisely which phenomena are merely correlated and which stand in a direct explanatory/relevance relationship to each other (cf. Mickelson forthcoming-b).³⁸ One thing that makes this abductive approach unique is that it explicitly treats indeterminism scenarios as a means for isolating and assessing the potentially freedom-undermining features found within normal determinism scenarios. In the context of this type of (in)dependent-variable generalization argument, the distinction between determinism scenarios and indeterminism scenarios is first and foremost a rhetorical tool; there is no presumption that this distinction tracks a freedom-relevant divide between possible universes.

Very roughly, here is how an (in)dependent-variable argument for the No-difference Premise might go. We could begin by selecting "the evolution of the physical world" as our independent variable, and take its values to be

37. The master generalization argument may be jarring at first, for it flies in the face of the common assumption that no single argument could cross the determinism/indeterminism divide; the standard strategy is to develop a constructive dilemma argument in which the case against compossibilism-friendly views and the case against libertarian-friendly views are developed independently (for serious worries about the constructive dilemma strategy, see Mickelson "Constructive Dilemma Arguments for the Impossibility of Free Will", ms.). Indeed, it might seem especially implausible that one could give a persuasive slippery-slope generalization argument in defense of Premise 2 of the master generalization argument. For one thing, it is evident from our discussion (above) that libertarian-friendly views of free will imply that there is a freedom-relevant difference between living in a deterministic world and an indeterministic one. It might seem, then, that the proponent of a libertarian-friendly view could block a relatively weak slippery-slope argument by appealing to their preferred answers to (E1) and (E2) of the explanation problem. However, the burden of proof created by the unstoppable generalization challenge is not so easily met. Yes, a *successful* libertarian-friendly answer to (E1) would constitute a generalization-blocking condition, but there are many rival libertarian-friendly solutions to the explanation problem and these candidate solutions correspond to different (and incompatible) replies to the unstoppable generalization challenge (cf. Mickelson forthcoming-b). As such, there is room for a lively in-house debate among proponents of libertarian-friendly accounts of free will about how to correctly respond to the unstoppable generalization challenge.

38. In theory, the abductive approach outlined here could be used to upgrade any argument incompossibilism to a solution to the explanation problem.

global deterministic causation (as described by determinism), global probabilistic causation, and global randomness.³⁹ By changing the value of this independent variable while keeping all the other details fixed, we can assess whether changes in the way the world evolves affect the value of the selected dependent variable, namely “one’s status as a free agent” (the values of which are “unfree” and “free”). For example, we might compare a case in which Vik lives in a deterministic world, a case in which Vik lives in a probabilistic world, and a case in which Vik lives in a world with no casual relations whatever.⁴⁰ If no change in the way Vik’s universe evolves corresponds to a change in Vik’s status, that is, no change in the causal relations changes Vik’s status from *unfree* to *free*, then we may abductively infer that evolution-related facts are not relevant to Vik’s lack of free will within the manipulation or determinism scenarios.⁴¹

Whether or not the abductive manipulation argument outlined above is persuasive, it reveals something very interesting and important about the basic dialectic of the free-will/determinism debate—something which will play a key role in my response to the problem of moral luck (below). In slogan form, it reveals that if indeterminism doesn’t “help” Vik to become a free agent, then we should abductively infer that it’s not the deterministic evolution (i.e., deterministic causation, deterministic laws, etc.) in the manipulation and determinism scenarios that “hurt.” More carefully, it seems that one must appeal to a libertarian-friendly solution to (E1) according to which free will is grounded in indeterminism of some kind in order to motivate an incompatibilist solution to (E2). This means that anyone who believes that libertarian-friendly solutions to (E1) of the explanation problem fail (i.e., none pinpoints a necessary condition on free will which is grounded in indeterminism) has no means to justify an incompatibilist solution to (E2).⁴² In short, impossibilists who reject all libertarian-friendly solutions to (E1) must find an *anti-incompatibilist* solution to (E2).

One obvious way to develop an anti-incompatibilist solution to the explanation problem is to draw upon one or both of the main noncausal sources of antecedent luck, namely circumstantial luck and/or constitutive luck. One might argue that antecedent circumstantial luck alone account for the lack of

39. As I argue in Mickelson (forthcoming-a, forthcoming-b), philosophers would benefit from using a more refined taxonomy for classifying the evolution of the world in this context (with categories based on the kind/amount of actual-sequence leeway in the evolution of the world), but the details of this taxonomy are not critical to the general point being made here.

40. That is, there are no *strong* laws of nature (Mickelson forthcoming-a, forthcoming-b, Perry 2004).

41. Nor does it show that the deterministic evolution in the determinism scenarios is completely irrelevant to Vik’s free will. For example, it is consistent with the conclusion that it is not the causal determination that makes Vik unfree that the deterministic relations in Vik’s world are actually *required* for Vik to satisfy certain necessary conditions on free will and moral responsibility (cf. Hobart 1934).

42. The criticisms of extant libertarian-friendly views are serious and well known (cf. Clarke 2003, Levy 2011, Pereboom 2014, van Inwagen 1983).

free will in determinism scenarios, but this does not seem very promising.⁴³ More plausibly, one might argue that it is the toxic combination of antecedent circumstantial and constitutive luck which destroys free will (cf. Latus 2001, Nagel 1986, 113–14). However, if constitutive luck is a freedom-underminer, one might reasonably hold that constitutive luck alone precludes free will—that is, given our lack of control over our constitutive properties, our lack of control over our circumstances and the causal relations which obtain in our world does not make us *worse off* with regards to free will and moral responsibility. Indeed, this is the solution to the problem of determinism that Strawson and other influential figures (e.g., Levy 2011) seem to prefer. Let us now turn our attention to a collection of arguments which are specifically designed to support this explanatory view.

4. THE PARADOX OF SELF-CREATION

The increasing popularity of source-condition solutions to (E1), the unmet-condition aspect of the explanation problem, reveals that many reflective people are drawn to the idea that a person must be a genuine author and owner of their actions in order to bear morally responsibility for them.⁴⁴ When we act, we do not merely want to *express a self* that has been bestowed upon us, that is, a self that is the mere product of factors beyond our control; we want our actions to express a self *of our own creation*. Although such a desire does not sound unreasonable on its face, one recurring worry in mainstream discussions of both free will and moral luck is that the type of robust self-creation or self-authorship that we intuitively want—indeed, believe ourselves to have—is metaphysically impossible. Worries of this kind fall under the umbrella the *problem of self-creation*.

Prior to philosophical reflection, it may seem that human actions spring *ab initio* from the mind of the actor himself. According to Roderick Chisholm, Aristotle expresses roughly this idea in his *Physics* with the statement: “Thus, a staff moves a stone, and is moved by a hand, which is moved by a man” (Chisholm 1964, 177). Here, claims Chisholm, Aristotle seems to be suggesting that there is something within the minds of normal human agents which allows them to initiate casual chains *ab initio*, such that human actors—unlike sticks, stones, and bones—are the ultimate authors and originators of their actions. Aristotle’s buck-stops-here account of human action is appealing, for it suggests that when we are called upon to account for human action, the

43. Carolina Sartorio (2015) comes close to suggesting this position with her discussion of backward-moving time travelers, for she suggests that their causal control over the past might allow them to be free even though we (non-time-travelers) are not.

44. It is interesting that Strawson has consistently embraced a source-condition solution to (E1) for the last few decades; meanwhile, libertarian-friendly possibilists have increasingly converged on the view that there is a source condition on free will, but differ from Strawson in their optimistic attempts to show that the sourcehood we intuitively want is grounded in indeterminism of some kind.

explanatory chains bottom out in precisely those private mental features which make humans agents rather than mere objects.

However, the plausibility of the Aristotelian story of human action quickly fades under scrutiny. When we try to make such generic descriptions of human action more precise, the paradox of self-creation emerges in one form or another. In her recent book *Aspiration*, Agnes Callard frames this paradox as a dilemma:

If, on the one hand, the value I cultivate in myself follows rationally from values I already have, then I do not do any creating. For in this case my “new” self was already contained in my old self. If, on the other hand, the new value is rationally unconnected to my earlier values, then its advent in my life cannot be my own doing. In this case, the self I end up with may be new, but it is not the product of my own agency. Rational value-cultivation is not *self-creation*, and nonrational value-cultivation is not *self-creation*. (Callard 2018; emphasis in original)

Here, Callard provides an argument for what is often expressed as a mere accusation. For example, it seems that Callard’s dilemma puts meat on the bones of Friedrich Nietzsche’s charge, levied in *Beyond Good and Evil*, that the notion of self-creation assumed in our commonsense notions of freedom and responsibility are incoherent:

The *causa sui* is the best self-contradiction that has yet been conceived, it is a sort of logical violation and unnaturalness; but the extravagant pride of man has managed to entangle itself profoundly and frightfully with this very folly. The desire for “freedom of will” in the superlative, metaphysical sense, such as still holds sway, unfortunately, in the minds of the half-educated, the desire to bear the entire and ultimate responsibility for one’s actions oneself, and to absolve God, the world, ancestors, chance, and society therefrom, involves nothing less than to be precisely this *causa sui*, and, with more than Munchausen daring, to pull oneself up into existence by the hair, out of the slough of nothingness. (Nietzsche 1886)

Unfortunately, most discussions of the paradox of self-creation are closer in rigor to Nietzsche’s than to Callard’s.

Callard is clear that her “dilemma” formulation of the paradox of self-creation builds upon the work of Strawson—the philosopher who has arguably done the most to establish the problem of self-creation as a philosophical problem in its own right. Strawson forwards his preferred characterization of the problem of self-creation under the label “The Basic Argument” (Strawson 1986, chap. 2; 1989, 1994, 2000, 2002, 2008, 2011). This name expresses Strawson’s view that the basic problem of freedom and moral responsibility is the paradox of self-creation, and it is the latter paradox that Strawson attempts to capture with his argument. Strawson has provided a variety of rigorously

stated premise-conclusion versions of the Basic Argument, but repeating just one of these careful formulations would invite quibbles over potentially insignificant details. Let us instead focus on the “core idea” behind the Basic Argument, which is nicely captured in the following summation of the argument from Michael McKenna and Derk Pereboom:

When an agent acts, she acts because of the way she is. But to be morally responsible for acting, the agent must then be morally responsible for the way she is, at least in key mental respects. But if an agent is to be morally responsible for the way she is in those key mental respects, she must be responsible for the way she is that resulted in those mental respects. This reasoning generates a regress, which indicates that finite beings like us can never satisfy the conditions on moral responsibility. (McKenna and Pereboom 2016, 265)

Unlike popular arguments in the free-will/determinism literature, such as manipulation arguments (discussed above) and the Consequence Argument (cf. van Inwagen 1983), the Basic Argument makes no mention of causal relations or the laws of nature. Rather, it draws our attention to antecedent factors beyond our control which are *internal* to the agent, factors which—despite being inside and, indeed, partly constitutive of the agent—still seem to be beyond the agent’s control in troubling ways. The aim of the Basic Argument, in other words, is to show us that, by our own intuitive standards of the control required for moral responsibility (i.e., the notion at play in the control principle), our lack of control over our constitutive properties is the fundamental threat to our free will and moral responsibility.

Although one need not state the paradox of self-creation as a regress argument (as Callard’s dilemma formulation demonstrates), Strawson’s regress argument is illuminating—far more so than its critics tend to admit.⁴⁵ Superficially, it may appear that the Basic Argument is offered as a knock-down argument for the impossibility and/or incoherence of free will (cf. Pereboom 2001). However, we might instead see the argument as generating a dialectical burden of proof on proponents of possibilism-friendly accounts of free will, that is, accounts of the control required for moral responsibility. Namely, it challenges possibilists to positively identify a principled way to stop the infinite regress; until a regress-blocker is positively identified, impossibilists have the dialectical upper-hand. This challenge, what we might call the *unstoppable regress challenge*, is roughly the same type of dialectical challenge imposed upon free-will possibilists by the unstoppable generalization challenge of the master generalization argument (above). However, the unstoppable regress challenge delivers a dialectical burden with its own special flavor. There seem to be at least two desiderata on an adequate reply to the unstoppable regress challenge:

45. There has been a promising uptick in explicit defenses of the Basic Argument over the last decade (cf. Coates 2017, Hartman 2018, Hendrickson 2007, Istvan 2011, Klemick 2013, Noggle 2008, Kane 2009, Marmorodoro and Mayr 2019, 171–72) which have pointed to serious flaws in standard criticisms of the Basic Argument (e.g., Clarke 2005, Fischer 2006, Pereboom 2001).

(i) it must not stop the regress in an ad hoc or otherwise arbitrary way (e.g., by insisting, without argument, that the regress in unproblematic or uninteresting⁴⁶), and (ii) it must not answer the regress in a way that gives rise to another vicious regress, where this new regress stopped in an ad hoc or arbitrary way (e.g., by insisting, without argument, that the new regress in not problematic and/or is not sufficiently interesting to warrant a reply⁴⁷). So, even if the unstoppable regress challenge does not decisively establish the impossibility of free will or moral responsibility, it gives us a distinct metric by which to measure the “cost” of possibilism-friendly accounts of free will.⁴⁸

The connections and differences between the unstoppable generalization challenge and the unstoppable regress challenge lend support not only to the proposal that these are interesting philosophical problems in their own right, but also suggest that a unified response must be given to them. This is worth noting, for it is well known that most of the leading possibilism-friendly solutions to the problem of determinism have elements which generate a seemingly vicious regress.⁴⁹ In the context of determinism scenarios, having a regress buried in the details of one’s preferred account of free will might seem like a small price to pay for a unified solution to standard arguments for incompatibilism and incompatibilism. However, such bullet-biting is more costly in

46. Strawson often remarks that his argument is frequently dismissed “as wrong, or irrelevant, or fatuous, or too rapid, or an expression of metaphysical megalomania” (Strawson 1994, 8). Such dismissive replies to the Basic Argument are dialectically infelicitous because they fail criterion (i).

47. For example, John Greco (1995, 95) attempts to emphasize the strength of his preferred solution to the paradox of moral luck by saying “Even moral worth is not completely isolated from luck, since it is not up to me what history I have or what restrictions exist on my personal identity. Nevertheless, moral worth is sufficiently isolated from luck so as to prevent any paradoxes from arising. And so our present solution does not solve one paradox only by making way for another.” However, as Latus points out, Greco does not show why it is a mistake to be troubled by the fact that we lack control over our constitutive properties; in effect, he has “merely asserted that there is nothing troubling about the fact that a person’s essential moral worth may be affected by luck” (Latus 2000, 169). If Latus is right, then Greco’s proposed solution fails criterion (ii). Along similar lines, Michael Moore (1997, 575–76) discusses a “character-stop manoeuvre” which might be used in response to the unstoppable regress challenge, but Moore provides reason to think that this manoeuvre fails criterion (ii).

48. Discussing his anti-incompatibilist “hard luck” impossibilism position, Levy notes that his view “has a clear predecessor in Galen Strawson’s well-known argument for the impossibility of moral responsibility,” but Levy suggests that his arguments are not vulnerable to standard criticisms of the Basic Argument (Levy 2011, 3–5). Arguably, though, Levy has misjudged the dialectical role of his arguments for hard-luck impossibilism. To the proponent of the Basic Argument, Levy’s case for hard-luck impossibilism serves as a *defense* of the Basic Argument. Specifically, it plays the dialectical role of showing that no possibilism-friendly account of free will delivers what it promises, that is, an account of “adequate control” which is both intuitively satisfying and coherent. As such, Levy’s arguments may be framed as a *rejoinder* to possibilism-friendly replies to the Basic Argument, showing why none successfully answers the unstoppable regress challenge.

49. For example, there is a “higher-order-desires” regress worry with Harry Frankfurt’s ahistorical compatibilism (cf. Frankfurt 1971, 21, Watson 1975) and regresses of different sorts plague extant libertarian-friendly accounts of free will as well (cf. Clarke 2003, Clarke and Capes 2017, Sec 3.2).

the dialectical context of the unstoppable regress challenge, for this challenge says that it is a success condition on an account of free will that it does not simply move the bump in the rug by trading one problematic regress for another. So, clarifying the distinct dialectic surrounding the unstoppable regress challenge makes it easier to see that arguments which address the paradox of self-creation cannot be dismissed as uninteresting or unimportant simply because one does not immediately feel the “intuitive pull” of the argument.

One reason that the unique and interesting contributions of the Basic Argument are easy to overlook is that the argument and its key upshots are often mischaracterized. For example, philosophers—including Strawson himself—often fail to adequately distinguish between the Basic Argument and summaries of the argument’s key upshots. For example, Strawson sometimes suggests that “the central idea” of the Basic Argument may be conveyed as follows:

1. Nothing can be *causa sui*—nothing can be the cause of itself.
2. In order to be truly morally responsible for one’s actions one would have to be *causa sui*, at least in certain crucial mental respects.
3. Therefore nothing can be truly morally responsible. (Strawson 1994, 5)

If we were to understand these three claims as constituting a free-standing argument, the argument would be so unpersuasive that one would be within their rights to dismiss it out of hand. There are, however, more charitable and fruitful ways to understand these three claims.

Claim 2 may plausibly be understood as a condensed statement of the *conclusion* of the Basic Argument. That is, Claim 2 is a quick way of saying that there is a necessary source condition on moral responsibility which can only be satisfied through an act of *ex nihilo* self-creation. This proposal may seem absurd on its face, but it is supported by the infinite regress challenge of the Basic Argument. That is, it seems that only an act of *ex nihilo* self-creation would constitute a genuinely regress-stopping action (where what qualifies as a “genuine” regress-stopping action is set by the two desiderata associated the unstoppable regress challenge noted above).⁵⁰ But, of course, the notion of a “nothing-self” implicit in the notion of a *causa sui* is incoherent; it is impossible “to pull oneself up into existence by the hair, out of the

50. Possibilists are welcome to respond to this attack upon their position, of course. However, if a possibilist contends that he has an adequate reply to the unstoppable regress challenge, he is obliged to frame it as such (i.e., he must demonstrate that it adequately satisfies the distinct dialectical burdens this challenge places on proponents of possibilism-friendly replies). This is serious challenge for the possibilists, given that familiar criticisms of possibilism-friendly accounts of free will pinpoint regress problems within these accounts (for the regress problem in Frankfurt’s compatibilism-friendly view, cf. Noggle 2008, McKenna and Pereboom 2016, 212–13; for regresses which plague extant libertarian-friendly accounts, cf. Clarke 2003). As such, it is striking that few possibilists take up this challenge directly. Robert Kane is perhaps the most notable exception: Kane says that his libertarian-friendly account of “self-forming actions” is also meant to be an account of “regress-stopping actions” (Kane 1996, 74; 2002, 277).

slough of nothingness.” Since it is impossible to be a *causa sui*, no possible action is a regress-stopping action. In short, since being a *causa sui* is impossible, it is impossible to satisfy the source condition on free will and, so impossible to satisfy the control condition on moral responsibility—which is just what Claims 1–3 above assert.

Another reason that the Basic Argument is widely misunderstood is that philosophers have an unclear view of the relationship between the dialectic of the self-creation paradox and the dialectic of the problem of free will and determinism scenarios.⁵¹ For example, the vast majority of philosophers classify the Basic Argument as an argument for incompatibilism, but this is untenable.⁵² The Basic Argument is an argument for incompatibilism when “incompatibilism” refers to mere impossibilism, but this is only because the spurious impossibility of free will and deterministic causal relations follows *a fortiori* from the impossibilist conclusion of the Basic Argument—just as the arbitrary “incompatibilist” view that free will is spuriously impossible with fluffy kittens follows *a fortiori* from the Basic Argument’s conclusion. What is more, the Basic Argument is an argument *against* incompatibilism when the term “incompatibilism” refers to the traditional incompatibilist view that deterministic causation (deterministic laws, etc.) destroys free will. In short, either way one disambiguates the term “incompatibilism,” this Procrustean label either misleadingly underdescribes or positively misdescribes the conclusion of the Basic Argument.

Notably, though, the correlation/explanation problem framework makes it easy to convey the upshot of the Basic Argument in a meaningful way

51. This is also evident in the popular classification of the Basic Argument as an “ultimacy argument for incompatibilism” or a “source-incompatibilist argument.” On this (outdated) taxonomy, the Basic Argument is in the same class as arguments for hard incompatibilism, roughly the view that deterministic evolution destroys free will in one way and indeterminism evolution destroys free will in another (cf. McKenna and Pereboom 2016, 149, Sartorio 2016, 148, n. 2). This taxonomy fails its users, for it misleadingly suggests that the Basic Argument delivers an incompatibilist response to (E2). However, the correlation/explanation framework makes it plain what these arguments do have in common. Namely, they support roughly the same kind of answer to (E1), that is, they all motivate a source-condition solution to (E1). On a better taxonomy, we would have a class of “ultimacy-condition” or “source-condition” arguments, and this class would include all arguments which propose a source-condition solution to (E1) (with subclasses including arguments for sourceimpossibilism, sourceincompatibilism, sourceimpossibilism, etc.) without implying that these arguments also deliver an incompatibilist solution to (E2).

52. To illustrate the worry, here is a concrete example: McKenna and Pereboom technically classify the Basic Argument as an argument for incompatibilism. In their recent introductory text on free will, the conclusion of the Basic Argument qualifies as an argument for incompatibilism on their proposed definition of the term (McKenna and Pereboom 2016, 30), yet they note that classifying the Basic Argument as an argument for traditional incompatibilism is problematic (2016, 151; see also McKenna 2010). Other philosophers have noted that there is something off about classifying the Basic Argument as an argument for traditional incompatibilism and have suggested their own preferred ways of handling this (cf. Levy 2011, Vihvelin 2013, 2018). However, the diagnosis offered here is coupled with the only proposed taxonomy of views which allows us to say, contrastively, exactly what the Basic Argument *does* and also *does not* purport to show. For further discussion of the need for an updated taxonomy of free-will views, see Mickelson (2015a, 2015b, forthcoming-a), and “Motte-and-Bailey Incompatibilism,” ms.

while still showing how it speaks to the problem of free will and determinism. First, the Basic Argument supports an impossibilist solution to the correlation problem. Second, it provides a source-condition solution to (E1) of the explanation problem. More carefully, it proposes that freedom-relevant sourcehood can only be achieved through an act of *ex nihilo* self-creation (as opposed, say, to the exercise of an actual-sequence ability to do otherwise or the exercise of metaphysically mysterious agent-causal powers). Third, the Basic Argument delivers a clear answer to (E2) of the explanation problem: the source condition described in (E1) goes unmet because the type of self-creation it demands is incoherent and/or metaphysically impossible. Notably, the truth value of determinism is irrelevant to whether someone self-creates *ex nihilo*, which helps us to understand why it follows from the conclusion of the Basic Argument that the traditional incompatibilist solution to (E2) is false.⁵³

In this section, I distinguished the problem of free will and determinism scenarios from the problem of self-creation, and then drew out new and interesting connections between them. In doing so, I have shown that the upshot of these ostensibly different problems is arguably the same, namely: it is impossible to coherently flesh out the necessary source condition on free will, that is, the source condition on the type of control required for moral responsibility. We are robbed of the type of sourcehood we intuitively want by the lack of control we have over our own constitutive properties, a.k.a. our constitutive luck. To be clear, this reveals that we can grant that the causal relations and non causal/agent-extrinsic circumstances which contribute to the production of our actions are entirely beyond our control without accepting that these factors play a role in *making* us unfree. With these lessons in hand, let us finally turn our attention to the paradox of moral luck.

5. THE PARADOX OF MORAL LUCK

In our everyday judgments of human actions, we assume that people are oftentimes morally responsible for what they do. However, once we grant this assumption, a review of particular cases of human action in which people are (by our commonsense intuitions) morally responsible for their actions reveals a troubling tension in our overall conception of moral responsibility. On the one hand, we are attracted to the idea that moral responsibility is immune luck. We do not, as Neil Levy points out, think it fitting to admonish someone for performing an immoral action by saying “you really ought to have been luckier” (Levy 2011, 54). This idea is captured by the control principle, roughly the thesis that one’s moral responsibility cannot be affected by factors beyond one’s control. On the other hand, when we look closely

53. For further explanation of the philosophical importance of recognizing that the conclusion of the Basic Argument implies the negation of incompatibilism, see Mickelson (2015a, 2015b, 2923–6; 2017).

at concrete case of human action, it seems that every action is due mostly, if not entirely, to factors beyond the actor's control (e.g., some combination of causal relations, external states of affairs, and the person's own basic personality and character traits). In short, our joint commitment to (i) the existence of morally responsible agents, (ii) the truth of the control principle, and (iii) our recognition that luck is a pervasive part of the human condition commits us to (iv) both accepting and rejecting the existence of moral luck, that is, cases in which factors beyond one's control affect (without undermining) a person's moral responsibility for his actions. This, in a nutshell, is the paradox of moral luck.⁵⁴

In his seminal work on moral luck, Thomas Nagel argues that our paradoxical stance on moral luck is not just a logical puzzle to be solved, or the unpalatable consequence of some controversial theory of moral responsibility. "The view that moral luck is paradoxical is not a *mistake*, ethical or logical," claims Nagel, "but a perception of one of the ways in which the intuitively acceptable conditions of moral judgment threaten to undermine it all" (Nagel 1976, 139). Put another way, Nagel seems to view apparent cases of moral luck as a rhetorical tool which helps us "to see how deep in the concept of responsibility the paradox is embedded" (Nagel 1976, 146).⁵⁵

If apparent cases of moral luck suggest that there is a problem buried *somewhere* in our commonsense views of morality and moral responsibility, it is reasonable to think that other rhetorical tools—such as determinism scenarios, manipulation scenarios, and self-creation puzzles—might help us to home in on the *exact location* of that problem. Indeed, it seems that the tools we have discussed in this paper point us to the same solution to the paradox of moral luck. Namely, when we start digging into our ordinary notion of moral responsibility through the hypothesis of determinism, we see that moral responsibility requires a certain type of control (i.e., the type of control at play in the control principle). Digging deeper, using tools such as the Basic Argument and manipulation arguments, we find that one must be a genuine source of one's actions to have this special type of control. However, when we attempt to positively state what this "genuine sourcehood" amounts to, we hit a wall: instead of finding a positive account of what sourcehood consists in, we find ourselves intuitively committed to the incoherent idea that sourcehood requires an act of *ex nihilo* self-creation. If this—call it the *source-paradox solution*—were right, it would explain why the problem of moral luck arises for us and why the problem seems so intractable.

Some readers may be suspicious of the source-paradox solution because it seems to be "imposed from outside" the mainstream moral-luck literature,

54. Nagel characterizes the paradox of moral luck as a follows: "A person can be morally responsible only for what he does; but what he does results from a great deal that he does not do; therefore he is not morally responsible for what he is and is not responsible for. (This is not a contradiction, but it is a paradox.)" (Nagel 1976, 146).

55. Bernard Williams seems to have roughly the same view, which is apparent when he says such things as "when I first introduced the expression *moral luck*, I expected to suggest an oxymoron" (Williams 1993, 251).

leaving the reader to wonder whether this solution really fits the unique dialectical contours of the problem of moral luck. Since Nagel is one of the founding fathers of the paradox of moral luck, let us close by assessing how well the source-paradox solution fares when it is cast as a solution to the paradox of moral luck as Nagel originally spelled out the problem.

To start, recall that Nagel says that he considers the “hypothesis of determinism” to be one among many ways of pushing people into the unnerving objective view of human action:

Something peculiar happens when we view action from an objective or external standpoint. Some of its most important features seem to vanish under the objective gaze. Actions seem no longer assignable to *individual sources*. ... The easiest way to produce this effect is to think of the possibility that all actions are causally determined, but it is not the only way (Nagel 1986, 110; my emphasis).

Along similar lines, Nagel tells us that the objective view helps us to see that “our capacity to view ourselves from outside—encourages our aspiration of autonomy by giving us the sense that we ought to be able to encompass ourselves completely, and thus become the *absolute source* of what we do” (Nagel 1986, 117–18; my emphasis). Translating Nagel’s claims into the language of the contemporary free-will literature, it seems that Nagel is open to accepting some type of source-condition solution to (E1) of the explanation problem of free will and determinism, and this is clearly consistent with my source-paradox solution to the problem of moral luck.

Although Nagel clearly recognizes the rhetorical value of the hypothesis of determinism, he is generally dismissive of the importance of the truth-value of determinism when it comes to the question of what *makes it the case* that people lack free will and, so, moral responsibility:

Once people are seen as parts of the world, *determined or not*, there seems no way to assign responsibility to them for what they do. ... [W]e can’t on close scrutiny make sense of the idea that what people do depends ultimately on them (Nagel 1986, 120; my emphasis).

In these passages, Nagel seems to be suggesting that there is a necessary source condition on free will (a.k.a. the metaphysical control required for moral responsibility) which is impossible to satisfy *whether determinism is true or not*. Moreover, when discussing standard views of free will, Nagel repeatedly indicates that both compatibilist and libertarian accounts of free will are inadequate—with compatibilist accounts being even worse than libertarian-friendly ones (cf. Nagel 1986, Ch. 7). As such, it seems that Nagel was open to accepting an impossibilist solution to the correlation problem, but was not keen on the view that libertarian-friendly accounts do enough to show how indeterminism “helps.” In the light of our discussion of the master generalization argument (in our above discussion of manipulation arguments), Nagel’s rejection of both compatibilist and libertarian-friendly accounts of

free will is not merely a negative, pessimistic view about the possibility of free will. If Nagel does not adopt a specific possibilism-friendly account of free will, he has no means to answer the unstoppable generalization challenge posed by the master manipulation argument and, as such, seems saddled with this argument's anti-incompatibilist conclusion. So, it seems that Nagel's commentary also fits well with the anti-incompatibilist solution to (E2) of the explanation problem that is proposed by the source-paradox solution.

Moreover, if Nagel does not accept any particular possibilism-friendly account of free will, he does not have a positive account of responsibility-relevant sourcehood which provides an answer to the unstoppable regress challenge of the Basic Argument. Perhaps it is unsurprising, then, that Nagel often focuses squarely on the problem of self-creation (though not under that label) when he laments what he sees while occupying the objective standpoint. For instance, Nagel suggests that one thing we can see from the objective view (via determinism scenarios, for instance), is that the very "objectivity that seems to offer greater control also reveals the *ultimate givenness of the self*" (Nagel 1986, 119; my emphasis). Elsewhere, Nagel suggests that the paradox of self-creation is *the* driving force in our dissatisfaction with possibilism-friendly accounts of the source condition on free will:

We cannot act from outside ourselves, nor create ourselves ex nihilo. But the impulse to this logically impossible goal also pushes us toward something else, which is not logically impossible and which may assuage the original impulse somewhat to the extent that we can attain it. (Nagel 1986, 126–27)

The objective view not only reveals to us our mistake in thinking that we are genuine self-creators, but this revelation leaves us with an insatiable longing for more sourcehood, authenticity, and ownership of our actions than anyone could possibly have:

I have to conclude that what we want is impossible, and that the desire for it is evoked precisely by the objective view of ourselves that reveals it to be impossible. At the moment when we see ourselves from outside as bits of the world, two things happen: we are no longer satisfied in action with anything less than intervention in the world from outside; and we see clearly that this makes no sense. (Nagel 1986, 117)

These two passages seem like an attempt to explain why possibilists—compatibilists and libertarians alike—find it so hard to abandon their quixotic project in the face of overwhelming evidence that their project cannot succeed.

Indeed, even when Nagel explicitly considers the relevance of actual-sequence alternative possibilities to free will, he says that the sort of actual-sequence alternatives that must be left open for free will would be closed off by constitutive luck *even if causal factors left actual-sequence possibilities open*:

While we cannot fully occupy this [external perspective] toward ourselves while acting, it seems possible that many of the alternatives that appear

to lie open when viewed from an internal perspective would seem closed from this outer point of view, if we could take it up. And even if some of them are left open, *given a complete specification of the condition of the agent and the circumstances of action*, it is not clear how this would leave anything further...that he could contribute *as source*, rather than merely as the scene of the outcome—the person whose act it is. If they are left open given everything about him, what does he have to do with result?” (Nagel 1986, 113–14; my emphasis)

Everything about the agent seems to be “swallowed up” by antecedent luck according to Nagel, “whether or not the relation between action and its antecedent conditions is considered as deterministic” (Nagel 1986, 114). In short, Nagel seems to be saying that it is irrelevant to one’s free will whether the causal relations of the world allow or rule out their actual-sequence ability to *do otherwise*; the fundamental threat to free will and moral responsibility is that no one has the actual-sequence ability to *be someone otherwise*.⁵⁶

Summing up, the source-paradox solution is clearly consistent with some of the most prominent ideological streams in Nagel’s thoughts about the problem of free will and the paradox of moral luck. We have seen that he is sympathetic towards a general source-condition solution to (E1). Moreover, Nagel’s repeated suggestion that part of what we want, as free moral beings, is to “create ourselves *ex nihilo*” and to escape the “ultimate givenness of the self” is consistent with the suggestion that the source condition requires an act of self-creation *ex nihilo*. If this is right, then the sourcehood we genuinely desire is impossible for us to manifest irrespective of whether determinism is true or false, and irrespective of what type of actual-sequence leeway is present in the evolution of the natural world. This implies an anti-incompatibilist, constitutive-luck solution to (E2). It hardly seems far-fetched, then, to propose that the source-paradox solution addresses the problem of moral luck as Nagel understood it.

To be clear, our review of Nagel’s writings has not shown that Nagel is committed to the source-paradox solution on the best interpretation of his writings, but nor was it meant to. Nagel’s writings on these matters is—despite being filled with useful insights and resonant phrases—too coarse-grained to be interpreted as an attempt to deliver one precise metaphysical explanation for and/or solution to the paradox of moral luck. Nagel clearly recognized that there were important distinctions and connections between the problem of moral luck, the problem of determinism, and the paradox of self-creation, but these distinctions and connections are not clearly identified in his writing.⁵⁷

56. For an interesting discussion of the incoherent desire to be someone else, see Arpaly (2006, 127–28).

57. Indeed, Nagel (1976, 1986) seems to cycle through the candidate solutions to (E2) discussed at the end of §2 (above), failing to note that these are rival explanations—to the same level of description—to the same problem. This strategy is rhetorically effective, insofar as it suggests that *one* of the many explanations has to be correct so pinpointing precisely which one is correct is “clean-up” work, which is perhaps why it is a go-to strategy among philosophers who write on free will for the general public (cf. Harris 2012).

Still, there seems to be an adequately good fit between Nagel's commentary and my source-paradox solution that it is plausible to contend that this solution addresses the problem of moral luck as it was originally conceived.

In closing, it is worth pointing out that the source-paradox solution explains why our intuitions commit us to the *paradox of moral luck*, but it also clearly entails the impossibility of moral responsibility and, so, the impossibility of *moral luck* itself. The source-paradox solution says that the paradox of moral luck arises because our ordinary notion of moral responsibility has a metaphysical control condition, and to satisfy this control condition, one must be the source of one's own action. To satisfy the source condition, however, requires the impossible; namely, an act of *ex nihilo* self-creation which would allow us to have control over our own constitutive properties. Since it is impossible for anyone to be the adequate source of their own actions, it is impossible for anyone to satisfy the control condition on moral responsibility, and, so, moral responsibility itself is impossible. So, if the source-paradox solution is correct, so too is the Skeptical View of moral luck, that is, the view that moral luck does not exist.⁵⁸

Arguments for the source-paradox solution are unlikely to uproot the stubborn commonsense intuitions which speak against it, so many people—myself included—will continue to experience paradoxical moral-luck intuitions. However, the fact that our intuitions can be resistant to rational argument is not a firm basis on which to build a critique of the source-paradox solution. In future work, we could strengthen the case for the source-paradox solution by arguing that resistance to the source paradox may be due to motivated reasoning. For instance, recent research in experimental philosophy suggests that people—and perhaps philosophers even more so (due the “paradox of expertise”)—respond in a biased way to arguments against the existence of free will and moral responsibility because they do not want to accept that conclusion (Clark et al. 2019).

Looking ahead, we could also flesh out the implications of the source-paradox solution. For example, assuming that the source-paradox solution is correct, an incoherent source condition is an integral part of our commonsense notions of free will and moral responsibility. It follows that there can be no coherent possibilism-friendly account of free will and moral responsibility, for the incoherence is built into the very data-point intuitions which a possibilism-friendly account of free will and moral responsibility would have to accommodate. As such, the source-paradox solution seems to imply that all possibilism-friendly accounts of free will—whether compatibilist or libertarian in spirit—are, *de facto*, the output of a revisionist methodology.⁵⁹ According to revisionists, our com-

58. The “cost” of rejecting basic desert and the moral practices based upon it has arguably gone down considerably in the light of recent work on the implications of “living without free will” (cf. Focquaert et al. 2019, Levy 2011, Milam 2016, Pereboom 2001, 2014, Waller 2011).

59. Moreover, if possibilists are engaged in *de facto* revisionism, then their lack of willingness to admit to this outright and deal with the consequences would make it reasonable to wonder whether they are also *de facto* proponents of Smilansky-style *illusionism*—for if they were illusionists, we could not be sure that they would admit it (cf. Smilansky 2003).

nonsense notions of free will and moral responsibility are seriously flawed. However, rather than seeing these flaws as grounds to abandon our commonsense commitment to the existence of free and morally responsible agents, the revisionist suggests that we should develop a possibilism-friendly account of free will and moral responsibility which departs from commonsense (cf. Vargas 2005, 2007). Of course, once we grant that the only viable possibilism-friendly project is to provide the best *revisionist* account of free will and moral responsibility, then all possibilism-friendly accounts of free will and moral responsibility must be judged by revisionist standards (cf. McCormick 2013).

6. CONCLUSION

In this paper, I have argued that there are distinct dialectics surrounding the problem of free will and determinism, the paradox of self-creation, and the paradox of moral luck, and have clarified the connections between them. Although I have framed my source-paradox solution as a reply to problem of moral luck, it constitutes a unified and dialectically felicitous response to all three problems. Whether a person is most drawn to the problem of free will and determinism, the paradox of self-creation, or the paradox of moral luck, my source-paradox solution says that the final metaphysical upshot of these inquiries is the same: the notion of control which backs our commonsense moral judgments (e.g., the notion of control in the control principle) is incoherent, so the commonsense view that we are sometimes morally responsible for our actions is false.

In the end, I expect that most readers will be unpersuaded by my arguments that the source-paradox solution to the problem of moral luck is correct or that the Skeptical View of moral luck is true. However, I have offered these readers new reasons to think that some central aspects of the narrative in the moral-luck literature are mistaken, and that correcting such problems might be worthwhile. Minimally, I hope this essay has shown that the standard “parallel-literatures” approach to answering the difficult question of what type of sourcehood is required for free will and moral responsibility is outdated, and a more rigorous “cross-pollination” approach may help us to better understand—and perhaps even resolve—the problems of free will, self-creation, and moral luck.⁶⁰

REFERENCES

- Mickelson, Mark B. 2014. Moral Luck as Moral Luck. *Southern Journal of Philosophy* 52, 1: 1–17.
- Nancy, Jean-Luc. 2006. *Being Singular Plural*. Translated by Connor B. Lawley. New York: Fordham University Press.

60. My thanks to the Swedish Research Council, the Department of Philosophy at the University of Gothenburg, and the Lund-Gothenburg Responsibility Project for funding. I would also like to thank Andrew Khoury and Robert Hartman for helpful comments on earlier versions of this paper.

- Callard, Agnes. 2018. *Aspiration: The Agency of Becoming*. Oxford: Oxford University Press.
- Campbell, Joseph. 2007. "Free Will and the Necessity of the Past." *Analysis* 67: 105–11.
- Chisholm, Roderick. 1964. "Human Freedom and the Self." In *Free Will*, ed. Derk Pereboom, 72–84. Indianapolis: Hackett Publishing Company Inc.
- Clark, Cory, Bo Winegard, and Roy Baumeister. 2019. "Forget the Folk: Moral Responsibility, Preservation Motives and Other Conditions for Compatibilism." *Frontiers in Psychology* 10. <https://doi.org/10.3389/fpsyg.2019.00214>.
- Clarke, Randolph. 2003. *Libertarian Accounts of Free Will*. Oxford, UK: Oxford University Press.
- . 2005. "On an Argument for the Impossibility of Moral Responsibility." *Midwest Studies in Philosophy* 29(1): 13–24.
- Clarke, Randolph and Justin Capes. 2017. "Incompatibilist (Nondeterministic) Theories of Free Will." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Spring. <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=incompatibilismtheories>.
- Capes, Justin. 2017. "The Basic Argument and Modest Moral Responsibility." *Analysis* 77(2): 156–70.
- Demetriou, Kristin M. (See also Kristin M. Mickelson). 2010. "The Soft-Line Solution to Pereboom's Four-Cases Argument." *Australasian Journal of Philosophy* 88(4): 595–611.
- Doohy, David, and Andrei Marmor. 2006. "The Case Against Moral Luck." *Law and Philosophy* 26(4): 405–36.
- Easton, John Martin. 2006. "The Cards That Are Dealt You." *Journal of Ethics* 10: 107–29.
- Foerquert, Parah, Gregg D. Caruso, Elizabeth Shaw, and Derk Pereboom. 2019. "Justice Without Retribution: Interdisciplinary Perspectives, Stakeholder Views and Practical Implications." *Neuroethics* 1–3. <https://doi.org/10.1007/s12152-019-09413-2>.
- Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66(2): 829–49. <https://doi.org/10.2307/2313811>.
- . 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68(1): 5–20.
- Gasco, John. 1995. "A Second Paradox Concerning Responsibility and Luck." *Metaphilosophy* 26(1–2): 81–96.
- Harris, Sam. 2012. *The Illusion of Free Will*. New York: Free Press.
- Hartman, Robert J. 2017. *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness*. New York: Routledge University Press.
- . 2018. "Constitutive Moral Luck and Strawson's Argument for the Impossibility of Moral Responsibility." *Journal of the American Philosophical Society* 4(2): 165–87.
- Hendrickson, Noel. 2007. "Improving the Metaphysical Argument Against Free Will." *Philosophical Papers* 36(2): 271–94.
- Hendova, Marcela. 2019. "Transformative Moral Luck." *Midwest Studies in Philosophy* 43(1): 162–180.
- Herdova, Marcela, and Stephen, Kearns. 2017. "This is a Tricky Situation: Situationism and Reason-Responsiveness." *The Journal of Ethics* 21(2): 131–54.
- Hobart, R. 1934. "Free Will As Involving Determination and Inconceivable Without It." *Mind* XLIII(169): 1–27.
- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford, UK: Oxford University Press.
- . 1990. "Logic and the Free Will Problem." *Social Theory and Practice* 16: 279–90.
- Istvan, Michael Anthony. 2011. "Concerning the Resilience of Galen Strawson's Basic Argument." *Philosophical Studies* 155(3): 399–420.
- James, William. 1884. "The Dilemma of Determinism," an address to the Harvard Divinity students, published in the Unitarian Review for September 1884; republished in *The Will to Believe and Other Essays*. London, UK: Longmans, Green, and Co.
- . 1890. *Principles of Psychology*, Vol. 1. New York: Henry Holt and Company.
- Kane, Robert. 1996. *The Significance of Free Will*. New York: Oxford University Press.
- . 2002. *The Oxford Handbook of Free Will*. New York: Oxford University Press.

- Nietzsche, Friedrich. 1886. *Beyond Good and Evil*, trans. Walter Kaufman. New York: Random House.
- Wolfe, Robert. 2008. "Autonomy and the Paradox of Self-Creation: Infinity Regress, Finite Selves, and the Limits of Authenticity." In *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy*, ed. James Stacey Taylor. Cambridge, UK: Cambridge University Press.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge, UK: Cambridge University Press.
- . 2014. *The Will: Agency and Meaning in Life*. New York: Oxford University Press.
- Petry, John. 2004. "Compatibilist Options." In *Freedom and Determinism*, ed. J. Campbell, M. O'Rourke and D. Shier, 231–54. Cambridge, UK: MIT Press.
- Recher, Nicholas. 1993. *Luck: The Illusion of Randomness in Everyday Life*. New York: Farrar, Straus, and Giroux.
- Sartono, Carolina. 2015. "The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will." In *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, ed. Alfred Mele, 255–73. Oxford, MS: Oxford University Press.
- . 2016. *Causation and Free Will*. Oxford, UK: Oxford University Press.
- . 2019. "Kinds of Moral Luck." In *The Routledge Handbook of the Philosophy and Psychology of Luck*, ed. Jan M. Church and Robert J. Hartman, 206–15. New York: Routledge.
- Samuels, Sam. 2003. "Free Will, Fundamental Dualism, and the Centrality of Illusion." *Kane* 2002: 489–505.
- Statman, Daniel. 2019. "The Definition of Luck and the Problem of Moral Luck." In *The Routledge Handbook of the Philosophy and Psychology of Luck*, ed. Jan M. Church and Robert J. Hartman, 195–205. New York: Routledge.
- Strawson, Galen. 1986. *Freedom and Belief*. Oxford, UK: Clarendon Press.
- . 1989. "Consciousness, Free Will, and the Unimportance of Determinism." *Inquiry* 32: 3–27.
- . 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75: 5–24.
- . 2000. "The Unhelpfulness of Indeterminism." *Philosophy and Phenomenological Research* 60: 149–55.
- . 2002. "The Bounds of Freedom." *Kane* 2002: 441–60.
- . 2008. "The impossibility of ultimate moral responsibility." In *Free Will: Second Edition*, 2009, ed. D. Pereboom, 289–306. Oxford, UK: Oxford University Press. Originally from *Real Materialism and Other Essays*.
- . 2011. "Free Will." In *Routledge Encyclopedia of Philosophy*, ed. E. Craig. London: Routledge. <<http://www.rep.routledge.com/article/Y014SECT33/pt.1>>. Originally archived 1998.
- Taylor, Richard. 1966. *Metaphysics*, 4th edn. Englewood Cliffs: Prentice-Hall.
- Tierney, Hannah. 2013. "A Manoeuvre Around the Modified Manipulation Argument." *Philosophical Studies* 165(3): 753–63.
- Todd, Patrick. 2011. "A New Approach to Manipulation Arguments." *Philosophical Studies* 162(1): 127–54.
- Tognazzini, Neal. 2011. "Owing Up to Luck." *Social Theory and Practice* 37(1): 95–112.
- Vargas, Manuel. 2005. "The Revisionist's Guide to Responsibility." *Philosophical Studies* 125(3): 309–420.
- . 2007. "Revisionism." In *Four Views on Free Will: Four Views on Free Will*, ed. John Martin Fischer, John Martin, Robert Kane, Derk Pereboom and Manuel Vargas, 166–167. Oxford: Blackwell.
- Vihvelin, Kadri. 2008. "Compatibilism, incompatibilism, and impossibilism." In *Contemporary Debates in Metaphysics*, ed. John Hawthorne, Theodore Sider and Dean Zimmerman, 103–18. Malden, MA: Blackwell Publishing.
- . 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press.

- . 2018. "Arguments for Incompatibilism." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Berlin: Springer. (Fall 2018 Edition). <<https://plato.stanford.edu/archives/fall2018/entries/incompatibilism-arguments/>>
- Valler, Bruce. 2011. *Against Moral Responsibility*. Cambridge, MA: MIT Press.
- Vatson, Gary. 1975. "Free Agency." *Journal of Philosophy* 72: 205–20.
- . 1993. "Postscript." In *Moral Luck*, ed. Daniel Statman. Albany: State University of New York Press.
- Williams, Bernard. 1976. "Moral Luck I." *Bernard Williams and Thomas Nagel 1976*: 13–32.
- Williams, Bernard, and Thomas Nagel. 1976. "Moral Luck." *Proceedings of the Aristotelian Society, Supplementary Volumes* 50: 115–135 + 137–151.
- Zimmerman, Michael J. 1987. "Luck and Moral Responsibility." *Ethics* 97(2): 204–26.