**Compatibilism as Non-Ideal Theory: A Manifesto[[1]](#footnote-1)**

**Robert H. Wallace**

Forthcoming in *Oxford Studies in Agency and Responsibility*, Volume 8, eds. David Shoemaker, Santiago Amaya, and Manuel Vargas. Oxford University Press. Please cite the published version.

**Abstract:**

This paper articulates and responds to a challenge to contemporary compatibilist views of free will. Despite the popularity and appeal of compatibilist theories, many are left with lingering doubts about compatibilism. This paper explains this doubt in terms of *the absurdity challenge*: because a compatibilist accepts that they do not have causal access to all the actual sufficient causal sources of their own agency, the compatibilist can find their own agency absurd. By taking a cue from political philosophy, this paper argues that a non-ideal construction of the problem of free will allows compatibilists to overcome this existential-metaphysical challenge, and by doing so, perhaps adopt a metaphysically progressive picture of human agency.

**Keywords:**

Free Will, Moral Responsibility, Compatibilism, Absurdity, Non-Ideal Theory, Metametaphysics

1. **Introduction**

In “The Impossibility of Moral Responsibility”, Galen Strawson reports that his regress argument against ultimate moral responsibility has convinced almost every student who has encountered it (1994: 8). The argument goes something like this:

1. A person is morally responsible only if they are the ultimate source of their own actions.
2. To be the ultimate source of their actions, a person would need to be the ultimate cause of who they are, at least with respect to some crucial aspects of their mental life and character.
3. But no one is the ultimate cause of who they are.
4. So, no one is ever truly morally responsible for their actions.

This is a seemingly powerful argument. Support for premise 3 comes in different varieties. Strawson, an impossibilist, suggests a vicious infinite regress of causation; our actions are brought about our character, which itself was brought about by prior actions, and those actions were brought about my character in the past, and that prior character was brought about by actions even further back…and so on. Responsibility is impossible. One could develop a hard determinist version of this argument, like Saul Smilanksy (2003), by suggesting that if determinism is true, then no one is the ultimate source of their actions, and so provide support for premise 3. As Smilanksy (2003: 275) argues, if determinism is true, then there is responsibility-undermining causal luck in the very fact that we are agents of a certain sort.

It has been my experience that undergraduates are quite taken with this style of argument. I certainly was when I was an undergraduate. But I am no longer enamored. Consider how Randolph Clarke (2005: 15-16, 20) and Michael McKenna (2008: 194) reply to Strawson’s impossibilism. The argument assumes that to be responsible for what you do, you must be responsible for who you are in crucial respects. But couldn’t one simply suggest that it is up to us—at least sometimes and in some important sense—whether we act on our existing beliefs, desires, and so on? The argument makes an implausibly strong assumption about the relationship between character and responsibility. Indeed, I agree with McKenna (2008: 194-197) and Carolina Sartorio (2014: 268-269, 2016: 148-146): ultimacy arguments involve a question-begging premise about the kind of control needed for moral responsibility. Any argument that assumes that free will requires ultimate control over the sources of one’s actions simply assumes the falsity of compatibilism. The compatibilist thesis just is that free will does not require causal control over all the actual sufficient sources of action, like the distant past and the laws of nature. It would take an independent argument to support the appeal to “ultimate” control, but this support would just be an argument against compatibilism. See? Nothing to worry about.

Yet when I offer these lines of reply to my undergraduates—whether they agree or not—they are left with lingering doubts about compatibilism. My students have noticed that compatibilism is *unsettling*, a fact that we professionals tend to forget. Surprisingly, it has been my experience that students sometimes find free will skepticism to be less unsettling than compatibilism even though this position involves comparatively greater revision to common sense.

And honestly, I get it. Even though I am a compatibilist, and I believe that disputants in the debate can all agree that compatibilist free agency is adequate for at least some interesting sense of free will and moral responsibility worth wanting, something feels wrongwith compatibilism. Compatibilists are naturalists about free will, arguing that free will is explicable in wholly natural terms compatible with deterministic physical law. At its best, compatibilism claims to present a “humble, unpretentious picture of human agency” (Sartorio 2014: 161). But I wonder if we have the right to claim such a picture. For what the compatibilist correctly says is real and worthy nevertheless leaves us anxious about our significance. If we accept the core of this naturalistic picture, what I will call the “givenness of agency”, the claim that our free actions and free agency are the determined product of causal forces extending beyond our causal reach, perhaps we should be *anxious naturalists,* as Simon Blackburn once suggested, philosophers “who fear that the end of the book will leave only melancholy” (1998: 160). This is a broadly existential-metaphysical cost to the view, a difficulty in making sense of life within a plausible metaphysics (Russell 2022: 520-521). My students wisely see that this is a cost to be taken seriously.

This paper is an attempt to earn that unpretentious picture of free agency without cost. I will proceed as follows. In section 2, I will articulate the cost of the unsettling feeling. I call it the absurdity challenge to compatibilism: the givenness of compatibilist agency renders it serious but apparently metaphysically insignificant, such that the compatibilist agent can find their own agency to be absurd. In section 3, I will discuss four ultimately unsatisfying strategies for responding to the challenge—debunking, patience, revisionism, and a combination thereof. To meet the absurdity challenge, I suggest in section 4 that we conceive of compatibilism as a research program in *non-ideal* metaphysical theorizing, a paradigm that, by analogy to non-ideal theory in political philosophy, rejects the problem of free will as constructed around problematically idealized assumptions. I detail some examples of potentially bad idealizations in section 5, and I suggest that they are problematic irrespective of whether one is a compatibilist or not. I conclude in section 6 that non-ideal compatibilism earns the right to a theory of free agency without metaphysical anxiety. In philosophy as in life we must lean into the anxiety to manage it.

1. **The Significance of Free Will and the Absurdity Challenge to Compatibilism**

In this section, I argue that compatibilism is unsettling because a compatibilist agent can find their situation absurd.

Like many, I like to start by saying that a theory of free will aims to explain the kind and degree of control needed to be morally responsible.[[2]](#footnote-2) Contemporary compatibilists like McKenna (2008) and Sartorio (2014, 2016) address ultimacy arguments within this framework. McKenna (2008: 199-200) has argued that it is perfectly felicitous to say that we are the “ultimate” sources of our own actions despite lacking more robust kinds of free will because we *already* accept that we do not need to be the ultimate sources of our actions to be morally responsible for them—imagine trying to downplay one’s bad idea by claiming that, although it was your idea, it ultimately originated with the big bang! “Ultimate” is a context-sensitive modifier, and, given the commitments of our existing practices, we need a metaphysically modest theory of ultimate responsibility, which McKenna believes can be provided by expressive considerations (2008: 198-199). Maybe our actions originate in us, per John Martin Fischer (2009: 10, 156), as they arise as contributions to the story of our lives. Or maybe, following Gary Watson (2008/1987) and P.F. Strawson (1962), our actions have salient emotional meaning to other agents, and this provides a direct sense of how we are the sources of meaningful action. Likewise, expressive considerations are important contributions to the meaning and value of a person’s life, sometimes anchoring a romantic sense of self in the world (Arpaly 2006: 40-43). This expressive kind of ultimate responsibility, as revealed in the quality of our reasons, cares, commitments, and values, makes our actions significant enough for day-to-day moral life. This is the kind of moral responsibility that seems reasonable to expect given that we are finite natural beings (McKenna 2008: 202-204).

At this point, I am convinced. This view expresses a profound commitment to an unpretentious picture of free agency and moral responsibility. But this is also the point where we might start to feel anxious. For the compatibilist agent canstep outside of this expressive and romantic perspective and see their life as a series of events unfolding with physical necessity under natural law. Indeed, that the compatibilist agent can do this just follows from the metaphysical humility of their view. By focusing on the importance of these expressive considerations, the compatibilist agent seems to face a kind of nauseating double-vision. They might think to themself: *I can’t help falling in love with you—*delightful and romantic and deeply meaningful—*and this is my only physically possible option*—not so much.

I worry that the framing of the free will debate solely in terms of moral responsibility makes it difficult to articulate this anxiety, since everyone should accept that the compatibilist view vindicates at least some interesting and important sense of moral responsibility. But as Robert Kane suggests, “debates about free will go beyond the practical concerns of law courts and everyday conflicts. They are about the human condition broadly conceived, about our place and importance in the scheme of things” (1998: 92). Even though compatibilists can offer plausible candidate theories of free will as the control needed to ground moral responsibility, theories that *I take to be successful*, they never seem satisfying. If the compatibilist is right, we are free and morally responsible *and yet* subject to disturbing influence by causes outside the scope of our own causal reach (cf. Russell 2017: 259). (You might think that this is because compatibilist theories fail to vindicate “basic” desert moral responsibility. I disagree. More on this in the section 5).

The problem is that a compatibilist agent can become awareof their disturbing metaphysical situation and so find their agency to be absurd. Here, I take a cue from Nagel (1971). He argued that life was absurd because we must make sincere efforts in our lives—we have life projects we cannot help but see as important—but can occupy an intellectual point of view, a “view from nowhere”, that reveals our efforts to be arbitrary. He writes that this absurdity arises from “the collision between the seriousness with which we take our lives and the perpetual possibility of regarding everything about which we are serious as arbitrary, or open to doubt.” (1971: 719). The compatibilist’s situation thus seems absurd in this Nagelian sense. The compatibilist takes their free will seriously, but they can also adopt a deterministic perspective from which that very freedom and responsibility is arbitrary or open to doubt by way of deep causal influence, by the fact that it is compatible with there being only one physically possible future. We can “zoom out” and consider the distant causal origins of our own agency. The problem is that compatibilist agency is serious but arbitrary, in that it grounds moral responsibility but is itself the product of forces far beyond the causal reach of free agents.[[3]](#footnote-3) Call this *the absurdity challenge* to compatibilism.[[4]](#footnote-4)

The absurdity challenge stems from a structural feature of compatibilism, namely, that it accepts what I call “the givenness of agency”. But let me clarify. Plausibly, any sensible view of human agency must accept that much about a person, including some of what constitutes their agency, is outside the scope of their causal reach. It is only at almost incoherent levels of self-creation that an agent might fully escape what is given about themselves, including aspects of their agency. The radical element of compatibilism is the totality of what is given, and this generates the absurdity problem.

A comparison with libertarian theories of free will is helpful here. Eleonore Stump suggests the following necessary condition for libertarian free will, understood in terms of the performance of a basic mental action that can be taken without performing any other actions: “a decision is free only if it is not the outcome of a causal chain that originates in a cause outside the agent” (2003: 302). Sloganized, this explanation of freedom is that *the buck stops with you*. But it is a structural featureof compatibilist views that they deny this libertarian necessary condition. For a compatibilist, free will is to be explained in terms of the right kind of causal antecedents, originating outside of the agent and far beyond the scope of an agent’s causal reach.[[5]](#footnote-5) In denying the libertarian’s necessary condition, the compatibilist seems to say that the buck does not stop with you. Yet the compatibilist has in hand a seemingly compelling theory of why you are to be held responsible for your actions. A theory, that is, of how your agency contributes to the worth of your actions, as to how your agential contribution makes certain modes of treatment fitting or appropriate or deserved: praise, blame, reward, and punishment. The absurdity is this: *that* you are the fitting or appropriate target of these modes of treatment in virtue of your free actions is, at least in part, beyond you.

If compatibilism is true, agents are free and morally responsible because they can aptly respond to a rich range of reasons for action (Fischer and Ravizza 1998, Nelkin 2011, McKenna 2013, Sartorio 2016), or because they have the right kind of mesh between their “real selves” and their actions (Frankfurt 1971, Watson 2014), or because they can do otherwise (Nelkin 2011, Vihvelin 2013), or because they have the right kind of moral emotional sensibilities (Wallace 1994, Russell 2017), or some combination of these, or perhaps something else. Each one of these features is to be taken seriously because they ground or partially ground the control needed for moral responsibility. In fact, I think they ground or partially ground aspects of free agency that are independently significant, like creativity, originality, and individuality. But the problem is that the very fact thatan agent has any of these features, and, indeed, the specific configuration and character of these features, is limited by what is physically possible under natural law given the past far beyond the agent’s causal reach. Being aware of this, a compatibilist can place their own free agency in a context where they can see its limits, and to see something’s limits just is to question its meaningfulness (Nozick 1981: 597).

Gary Watson (2008/1987) has forcefully expressed this awareness of limits in “Responsibility and the Limits of Evil”. When we reflect on formative tragedy, on how moral innocents become evildoers through trauma and suffering, Watson says that we feel an “ontological shudder”, a kind of recognition that we too could have been evil given different causal pressures (2008/1987: 132). We tend to think, as my grandmother used to say, “there but for the grace of God go I”. And once you feel the shudder in the bad case, you can come to feel it in every case, because it is not as if only evil people are subject to formative causes beyond their reach. We all are. Of course, there are limits to our modal fragility. We cannot be evil *as we are now*; the possibility of our being evil is precluded by the actual causal antecedents that have shaped us, including our own choices (McKenna 2008/1998: 216). But the fact that we cannot be evil given our actual causal history does not address the absurdity challenge. Our ability to see ourselves as an object of causal pressure such that we *could* *have been different* “without developing the illusion that we are able to escape from [our] highly specific and idiosyncratic position” is itself the source of the issue (Nagel 1971: 720).

In this section, I raised the absurdity challenge to compatibilism: compatibilists can feel an absurd disparity between the seriousness with which we must take our free agency and the larger naturalistic perspective from which that very agency seems arbitrary. The challenge does not defeat compatibilist theories of free will, but instead, to steal phrase from Nagel (1971: 724), lends these theories a “peculiar flavor”.

1. **Potential Replies**

Perhaps the absurdity challenge can be met with existing compatibilist resources. I am skeptical. In this section, I consider and reject four replies to absurdity challenge.

Some compatibilists might optimistically dismiss the absurdity challenge. Perhaps the appearance of absurdity is just the product of what Daniel Dennett (1984) calls “bugbears” and “bogeymen”, incompatibilist thought experiments or metaphorical framing devices that prime us to accept incompatibilist conclusions without close inspection. Indeed, I have intentionally done what Dennett says philosophers should not, namely, “fanned the coals of our anxieties” about “the human condition” (1984: 4). Inspired by Dennettian resources, a debunking response to the absurdity challenge might say that the “peculiar flavor” of compatibilism is a product of some imaginary or metaphorical threat, and so the compatibilist should not find their situation absurd.

Yet the absurdity is not the product of any of Dennett’s well-known bugbears. There is no trick about which we can become disillusioned. We do not imagine an invisible jailer keeping us in place. We are not *mere* playthings of the cosmos. No one can know what we are going to do before we do it. We are not decomposable into mindless mechanisms. And there is no “dread secret” behind the free will debate that, if discovered, would shock us to our core (Dennett 1984: 7-17). It is just that, free and responsible as I am, I can also step outside of the view of my own free agency and see my very agency as the product of forces far beyond my own causal reach. I cannot see how this is a metaphor priming me to reject compatibilism. It is something that I can do. Dennett is right, of course, that the bugbears can infelicitously support incompatibilist conclusions, for they prime us to accept question-begging assumptions, presumably about “ultimate” responsibility. Nevertheless, a Dennett-style diagnosis of the absurdity challenges misses the point. As Russell (2017: 250-255) has convincingly argued, “bugbear-busting” fails to address the emotional resonance of anti-compatibilist intuitions.

Maybe we should be pessimists about compatibilism and patiently accept that the compatibilist agent will find their situation absurd. Here, I take inspiration from Paul Russell (2017). Russell suggests that compatibilism is unsettling because our ethical thinking has internalized the false assumption that genuine moral responsibility is incompatible with fate and luck, an assumption which is the product of historical and contingent commitments about the self and the world (2017: 256). A mild despair at our metaphysical situation seems appropriate (2017: 258). Yet Russell (2017: 258) suggests that our negative feelings concerning compatibilism are “likely to recede or dissipate over time as we (moderns, Westerners) withdraw from” what Bernard Williams (1985) calls “the morality system”—the system of deontic morality that precludes luck and fate.

I agree that the absurdity challenge arises for us (contemporary Westerners) in the context of historically contingent commitments about deontic morality and autonomous individuality. But we cannot resolve the problem by rejecting these commitments or waiting for them to change. For instance, the compatibilist can face the absurdity challenge cast in terms of virtue even though virtue attributions are widely held to be compatibilist friendly. I might admire the human excellence of a courageous person but can take up a perspective in which their excellence is more like the natural beauty of a snowcapped mountain—a part of the natural unfolding of things governed by physical law—rather than a chosen feature of their character that is an ultimate expression of who they are. This will not diminish the admirability of their courage, but instead, make the *agency* in their courage less insignificant. In fact, any domain in which we might evaluate someone’s agentic contributions will cause trouble for the compatibilist. Consider, for instance, Salieri’s jealously of Mozart’s genius in the film *Amadeus*; God, the causal source of all things, seems to mock Salieri’s chosen discipline by gifting genius to Mozart, a deeply unserious man. As Kane has put it, we naturally wonder about the “objective worth” of our actions, whether they “ultimately redound to us or not” (1998: 97-98). In principle, questions about our significance extend to the many ways that human persons come to assess the worth of their actions and projects.

Perhaps, then, we need to revise widespread concepts about agency to avoid the absurdity challenge. Here, I draw inspiration from Manuel Vargas (2013: 73). Vargas argues that our ordinary concept of free will involves libertarian components which are unlikely to be true. He suggests that we can justify our responsibility practices because they cultivate agency of a valuable sort, agency that is sensitive to moral considerations within a set of cultural and social circumstances (2013: 213-214). We could try to design social-ethical practices that cultivate the valuable, naturalistic kinds of agency (2013: 196-198). As a response to the absurdity challenge, one might say that a naturalistically plausible and revisionary account of our moral responsibility practices will allow us cut away whatever problematic sense of metaphysical significance generates the absurdity challenge, leaving only what is worth wanting, namely, the agency needed to be morally responsible.

While revisionism looks promising, the revisionary compatibilist can still occupy a position from which their revised conception of free agency seems arbitrary or open to doubt. Reflecting on the fact that we *can* cultivate agency of a certain sort by way of social-ethical practices invites the absurdity, for this fact about ourselves serves to highlight our ontological fragility in the causal nexus. The looming anxiety about significance does not merely stem from distant causal sources, but more local cultural, political, and economic outside of our immediate causal reach. The compatibilist might think to themself, for instance: *I can aptly respond to moral considerations, and this is of great significance*—but they can also think this—*and if I had been born in different circumstances, I would not have been, and maybe could not have been, sensitive to these considerations at all*. Revisionism cannot cut off the absurdity challenge at its root because it too is committed to the givenness of agency.

A fourth response to the absurdity challenge might draw on both genealogical and revisionist considerations. Maybe the challenge depends on *uniquely* contemporary and Western commitments, such that the resources needed to suitably revise our concepts, or even debunk the threat of absurdity altogether, lie outside of our local historical, cultural, and political context.[[6]](#footnote-6)

I am sympathetic to this idea. By reacting to ever more insidious naturalistic threats to freedom, we have developed an intuitive sense of self as agent that is highly idealized and metaphysical (Kane 1998: 91-96). I say that historical and comparative work can help us come to grips with the reality of our metaphysical situation, for many views are committed to the givenness of agency, if not by way of a commitment to naturalism (Wallace 2021: 329-331). But as such, this work cannot furnish us with concepts ready-made for successful revisionary conceptual engineering against the absurdity challenge. These views face their version of the challenge too.

For instance, Augustine maintains the (arguably) compatibilist view that we are responsible for what we do even though the grounds of our agency lie beyond our control, for we only have a properly functioning free will through God’s unmerited grace (Baker 2003). Herein lies a problem. The view seems to make God ultimately responsible for how we use our free will (Rogers 2004, Stump 2014: 117). Perhaps this is an ultimacy objection that should be rejected. But notice in any case how it exposes the Augustinian compatibilist to the threat of absurdity. They must treat their free agency seriously—eternal damnation is at stake! —but can see themselves from within a broader theological perspective from which that same agency seems (literally) arbitrary. Indeed, Calvinist theological determinism may seem psychologically preferable to the “anxious theism” implied by this view.

Here's another example. An (arguably) compatibilist project in Buddhist philosophy aims to square moral responsibility (and so perhaps free will) with the fundamental metaphysics of dependent origination, the view that everything is conditioned (or caused) such that nothing substantially exists independently of anything else (Finnigan 2022: 9). Dependent origination seems to threaten moral responsibility, since it threatens the substantial existence of the self. But perhaps it could be conventionally or instrumentally true that persons are morally responsible for their actions even if, ultimately, there are no persons at all (Sideret 2017, Finnigan 2022: 11-12, 16-17). Nāgārjuna-style skepticism offers a powerful objection to construing this position in a realist way: at best, what is conventionally or instrumentally true is ultimately just useful rhetoric (Gier and Kjellberg 2004: 278). Now, even if this objection fails, it highlights the fact that the ultimate and conventional truths do not match, and yet we are bound to take the conventional seriously. This renders us apparently absurd.

Given the foregoing, I remain unconvinced that existing resources, even when combined, are adequate in meeting the absurdity challenge. Indeed, although my focus in this paper is on the naturalistic commitments of compatibilists in the contemporary free will debate, the threat of the absurdity challenge is pervasive. All in all, compatibilists will have to try something new.

1. **Towards a Non-Ideal Theory of Free Will**

The absurdity challenge suggests that compatibilist agency is absurd—serious but insignificant. I think that this challenge can be diffused by interrogating the way we set up the problem of free will in the first place.

To see why, let’s consider Nagel himself, for he is an anxious naturalist *par excellence*. In *The View from Nowhere* (1971), Nagel gets close to articulating the absurdity challenge. He writes that “it is impossible to give a coherent account of the internal view of action which is under threat” by determinism and naturalism, and so all accounts of free will “fail to allay the feeling that, looked at from far enough outside, agents are helpless.” (1971:113). Once we see ourselves from the outside, we become totally dissatisfied by anything less than an ability to “encompass ourselves completely” (1971: 119). And this is ironic, because our agency is given to us, but our means of becoming freer and so more responsible is to take a step back from this givenness, to subject our beliefs and desires to critical examination. The process “that starts as a means to the enlargement of freedom seems to lead to its destruction” (1971: 120). Nagel ultimately recommends that we adopt an *incomplete objective view* that leaves a blind spot where our agency is concerned and stop investigating our agency, lest we undermine it. The best we can do is to hope that our actions would not be unacceptable from a more complete objective view (1971: 126-130). Nagel thinks he has hoisted the compatibilist by their own petard by showing how compatibilist-friendly resources are self-undermining.[[7]](#footnote-7)

Yet Nagel has made a mistake. The “view from nowhere” is an abstraction. Really, it is an *idealization*, the ideal (and imaginary) endpoint of our ability to step outside of our subjective viewpoint. Nagel’s claim here is that this ideally external and objective viewpoint does not only, as a descriptive matter, leave us unsatisfied with nothing less than metaphysical abundance, the assumption here is that this viewpoint *should* leave us unsatisfied. And I worry that the assumption that we should be metaphysical dissatisfied by anything less than metaphysical abundance leads to a way of setting up what counts as a successful theory of free will that always burdens compatibilism with the existential-metaphysical cost of the absurdity challenge. This set-up unfairly validates our uneasy feelings about the view. So, what the compatibilist needs is a critique of this set-up. I suggest that we frame compatibilism about free will as a *non-ideal theory* in metaphysics, which rejects setting up the problem in terms of an idealized viewpoint.

Hopefully that sounded provocative. I mean it as a technical point. The term “ideal theory” stems from Rawls (1971), but the ideal/non-ideal theory distinction has come to mean many things in political philosophy (Valentini 2012: 654-655). As L.A. Paul (2012: 11) notes, all theories idealize to some degree—they are plausibly abstract representational structures that require some idealization to clearly distinguish between the model and the world. And this point goes for theories in metaphysics too. The use of neuron diagrams in the causation literature or the employment of thought experiments concerning questions of composition requires idealized abstractions away from the messiness of actuality (Paul 2012: 12).

I have no problem in general with idealized assumptions in metaphysics. I will myself suggest a way that ideality could be helpful in section 6. The sense of “non-ideal” I employ focuses on ways idealization can go badly. I have two such ways in mind.

First, some problematic idealizations occur in *problem construction*. As David Schmidtz (2011: 775-776) has pointed out, some theories employ idealized assumptions about the very problem that they are designed to solve; an ideal way to solve traffic flow might end up creating confused and angry drivers. Part of Rawls’ (1971) ideal theory famously assumes that agents will strictly comply with principles of justice, but as Schmidtz (2011: 778) argues, when you choose a set of principles, you also choose your compliance problem—how to get actual agents to comply with those standards. Thus, choosing principles of justice is also a matter of choosing what kind of compliance problem you want to live with. So, you cannot simply idealize compliance away in your theory of justice insofar as a theory of justice aims to explain how justice could be actualized.

Second, some theories involve models that are so idealized that they become unhelpfulin terms of *actual guidance*. Charles Mills (2005: 168-169) argues that ideal theory in prominent strains of political philosophy involves idealized models of the phenomena it aims to explain that are ultimately unhelpful or even harmful; an idealized theory of justice might abstract away from relations of social domination, require idealized notions of the capacities of the agents represented by the model, and by and large ignore the impact of oppression—for instance, on the capacities of the oppressed. An idealized theory of this sort might produce bad results about the very problem the theory purports to deal with because it obscures the conditions in which the theory is intended to be implemented.

By analogy to non-ideal theorists in political philosophy, compatibilists might reject idealized misconstructions of the problem of free will. How will this help with the absurdity challenge? These idealizations, or really, an idealizing perspective, are the source of the apparent absurdity. We appear absurd to ourselves only because we make the deterministic or natural perspective as one from which our freedom, responsibility, and significance needs shielding. My suggestion is that the compatibilist radically accept their metaphysical situation. They should insist that constructions of the free will debate that render this acceptance normatively suspect—constructions that accept that we *should* find compatibilism, not merely that we do—make a mistake at the level of problem construction and in theory guidance.

Having explicated the sense of non-ideal theory I aim to employ, in the next section, I consider examples of problematic idealizations and use them to offer a partial characterization of the problem with an “ideal” metaphysics of free will. I will then return to the absurdity challenge in section 6.

1. **Spotting Problematic Idealizations**

I now consider three interrelated examples of problematic idealization: the significance of free will considered by itself, manipulation arguments, and the notion of “basic” desert. My aim is to sketch out how a non-ideal theorist of agency might spot problematic idealizations in the current debate about free will and then show how these idealizations unfairly validate the feeling that compatibilist agency is absurd.

Consider first that the problem of giving an account of free will’s significance is itself sometimes framed in idealized notions. That our plans and projects, and indeed, our very selves, are worthy of respect seems to follow from the fact that “the buck stops with” us, that we are the sorts of beings who can be *truly* unique and individual (Kane 1998: 87). Buck-stopping intuitions abound when we consider the possible sources of free will’s significance. Determinism apparently renders genuine creativity suspect, as what we create is implicitly contained in the natural causal antecedents of our actions (Kane 1996: 81), and it appears to rob us of our autonomy by subjecting us to laws beyond ourselves (Kane 1996: 82). Determinism threatens our uniqueness, insofar as what explains our actions and life projects is not exclusive to us but could be replicated by someone else with a similar causal history (Kane 1998: 85-86); think again of Watson’s shudder, that if I had been subject to different causal pressures, I would have been a different sort of person. Our excitement about life, our sense of our own life-hopes, would be altered given the realization that we contribute to them in a way consistent with physical determinism (1998: 97-88).

I have two related concerns about this. First, recall that any sensible theory of free agency must say that some of who we are, including some our agency, is the product of antecedent conditions over which we have no control. Grand talk of *true* and *real* autonomy, creativity, and individuality obscures this. Second, given what I have argued earlier, explaining the significance of free will in terms of buck-stopping can easily slip into the question-begging assumption that there could be no successful compatibilist account of creativity, autonomy, uniqueness, or respect. We might wrongly assume that since the givenness of agency seems to render these aspects of ourselves contingent and modally fragile, no compatibilist account of these features could succeed.

Our tendency to idealize free will’s significance is related to another idealization that often appears in the context of manipulation arguments. Manipulation arguments have the following general structure:

1. If an agent is manipulated in some manner *M* to perform any action *A* from whatever the compatibilist takes to be minimally sufficient for freely acting, then that agent is not free and so not responsible for *A*-ing.
2. Any agent manipulated in manner *M* to *A* is no different than any normal agent who is physically determined to *A*.
3. So, physically determined normal agents are not free and responsible for their actions.

All the action in this style of argument occurs in the second premise, which is normally defended by appeal to cases. We imagine a person who is genetically modified before birth (Mele 2006: 188-195) or who was manipulated by neuroscientists or is indoctrinated by their culture (Pereboom 2014: 74-82). We intuitively feel that such a person is not a morally responsible agent.

But notice that these cases are designed to *only* highlight the contingency and fragility that we might wish to theorize away when we consider free will’s significance by itself. As Nomy Arpaly (2006: 110-117) has pointed out, we often feel differently when we consider real life cases of religious or political conversion or traumatic accidents. People *actually* undergo radical change while having intact agential capacities, as in our thought experiments, and yet we often do not have intuitions of non-responsibility in these real-life cases. Rather than seeing real life cases as responsibility-undermining “manipulations”, we see them for what they are, namely, examples of the sort of contingency and fragility that characterizes ordinary human agency from a naturalistic point of view. Thus, a non-ideal theorist might assert that rather than employing a thought experiment to help us explain a part of actuality, the use of the manipulation cases idealizes away the very actuality it is intended to help us explain. Thus, we have an issue in problem construction.

Now, one could say, as Dana Nelkin (2013: 125) does, that the unrealistic quality of the manipulation cases “helps ensure that we are focused on the stipulated features, and that we aren’t implicitly but unconsciously relying on background assumptions that we bring to ordinary life.” We could set up a debate here between an ideal and non-ideal theorist about agency about the status of these cases. But the claim that the manipulation cases help us get away from problematic commonsense intuitions is precisely the sort of thing a non-ideal compatibilist ought to deny. The abstract condition seems to me to be the one where problematic background assumptions about the significance of free will are likely to come to the fore as we anxiously back away from the givenness of our agency. But, of course, the compatibilist view is committed to the givenness of agency, and the actual cases of radical personal transformation show us that common sense is not wholly disturbed by at least some givenness in ordinary circumstances. Thus, the idealizations—the one at the level of problem construction concerning the significance of free will and the other at the level of appeal to intuition in abstract manipulation cases—are related.

Manipulation arguments are often fixed on desert *in the basic sense*, desert of praise and blame not on the basis of consequentialist or contractualist reasons (of the sort that might be attractive to compatibilists) but solely because the agent performed the action (Pereboom 2014: 2, Sartorio 2016: 7). Basic desert is “not justified in terms of any more fundamental normative consideration” (McKenna 2019: 246). The basic desert framing is one way of answering the question about the significance of free will. It tells us why free will is morally significant, namely, that it preserves “real” desert. I am somewhat confident that compatibilists can account for basic desert. Nevertheless, and at the risk of sounding cavalier, let me suggest that focusing on basic desert has the marks of a problematic idealization because it leads to bad problem construction and undermines actual guidance.

Consider issues of problem construction. Basic desert is not a feature of many plausible moral theories that employ desert concepts (Vargas 2013: 257-258; Kristjánsson 2003: 55-60), and it is even arguable that basic desert is the product of an overly demanding theory of morality, per Russell (2017: 254-261). This leads to odd consequences in problem construction. It is an open question which of the extant normative ethical theories is correct, and it would be very strange if the moral significance of the free will debate requires that consequentialist or contractualist principles not be normatively fundamental, or that no interesting question about the metaphysics of control or agency arises if they are. I have already tacitly expressed skepticism about this narrow view of the problem of free will and determinism by offering examples related to the worth of virtue and artistic talent. And yet the basic desert framing seems to entail that no interesting questions about free will as such can feature outside of theories that feature normatively fundamental desert. Indeed, many compatibilists are happy to accept the view that no one deserves praise and blame in the basic desert sense because they have commitments to moral theories that do not feature basic desert.[[8]](#footnote-8) So, I think that constructing the free will debate around basic desert moral responsibility is a problematic idealization because it allows us to ignore whole sets of plausible views in order to neatly frame our metaphysical question.

Moving on to issues of guidance, since basic desert might not even be central to our responsibility practices as we find them, it might be actively unhelpful in theorizing about our actual practices, either descriptively or normatively (Wallace 1994: 106-109, Shoemaker 2015: 106-107 and 2018: 78). While it is true that many have basic desert intuitions, it is simply not clear that ordinary thought and practice includes basic desert presuppositions, or if they do, to what degree. Without a clear picture of the commitments of our actual practices and psychology, arguments for revising these practices on basic desert grounds might be actively unhelpful. One worries, for instance, as to how metaphysically driven revisions of our angry moral responsibility practices, as in Pereboom (2014, 2017, 2021), can adequately speak to the injustice faced by the oppressed in facing affective injustice, a conflict between expressing apt or fitting anger or repressing this anger as counterproductive (Srinivassan 2018: 13-15). Conversely, if we take basic desert to be vindicated by an account of free will, be it libertarian or compatibilist, we might inadvertently fail to recognize the merit of alternative, non-desert modes of response to wrongdoing (McKenna 2021). A less idealized conception of the free will problem could help us avoid bad recommendations about revising or keeping existing practices.

These three examples have gone by too quickly to convince. But in spotting them as idealizations my goal was not to convince but rather exemplify. What do all these idealizations have in common? Metaphysics is largely a matter of theory choice by way of theoretical virtues, among other considerations (cf. Sider: 2009: 385). Devastating objections are rare, and so theory choice is largely settled by weighing the theoretical costs and benefits of a view. My suggestion is that each idealization involves a theoretical vice analogous to *empirical inadequacy*. Empirical inadequacy happens when a theory does not explain the empirical data. But metaphysics too has its data: ordinary experience as defeasible evidence about the nature of the world, thought experiments, and our arguments (Paul 2012: 16, 20, 21-22). The problem of free will arises from our need to understand the human condition in the context of our general metaphysical theories, and so that condition is a core piece of the data. A theory that mischaracterizes the human condition distorts the data, and so such a theory is not adequate to the phenomena it aims to explain. A mischaracterized condition thus produces bad constructions of the problem that raise concerns about actual guidance. A non-ideal compatibilist should assert that their theory only bears the existential-metaphysical cost of the absurdity challenge within this problematic construction of the free will debate.

More on that in the next section. But first, I need to respond to two objections. The first objection says that I have just given examples of how incompatibilists beg the questions against compatibilists. But many compatibilists *accept* the idealized terms of the debate which I am attempting to characterize, so I do not think this objection is on the right track. My position opposes theirs. The second objection says that my way of construing these examples as bad idealizations presupposes that compatibilism is true, and so I beg the question against incompatibilists. But my point is not that we all must accept the compatibilist’s radical view about the givenness agency. Rather, my point is that we have constructed our problem space in a way that stacks the deck against accepting (or sometimes even acknowledging the possibility of) the givenness of agency. Both objections fail to see that these problematic idealizations are problematic *independent* of compatibilism. I fear that they reflect an unreasonable, idealized sense of the human metaphysical condition, one that *leads to* question-begging assumptions regarding compatibilism. I have been writing for other compatibilists because I think they are in need of a non-ideal theory, but down-to-earth libertarians and free will skeptics should welcome a non-ideal metaphysics of agency too.

1. **Once More with an Unsettling Feeling**

Having argued for a conception of a non-ideal metaphysical approach to free will, I now argue that this approach can defuse the threat of the absurdity challenge, thus allaying the existential-metaphysical cost of compatibilism.

I say that the non-ideal compatibilist can allay this cost by rejecting constructions of “the problem of free will and determinism” as organized around ideals that render the acceptance of the givenness of agency normatively suspect; thus, the absurdity challenge fails because the theoretical presuppositions that underpin it have been rejected.

Let me explain by comparison. Susan Wolf (2015) has argued that there is a kind of mistake in inferring from the fact that there is no “Meaning of Life” to the conclusion that life has no meaning. The desire for an ultimate or final purpose does not render the project of endlessly playing sudoku as meaningless as a life spent in service to others. We do have a sense that some aims are more meaningful than others, and Wolf explains this in terms of a confluence: finding subjective fulfillment in doing objectively valuable projects. The absurdity of our lives, Wolf suggests, should be rejected. To the person unsettled by their cosmic insignificance, she says: “the only advice I can give to such a person is: Get Over It.” (2015: 104). Of course, by this point she has a *theory* in virtue of which she can say this, and I take that theory to include a critique of a construction of “the problem of life’s meaning” in idealized terms such that her commonsense solution is renderedinadequate. Someone who accepts her view of the problem no longer needs to feel like their life is absurd. Likewise, I say that we should not think that compatibilism is absurd, contrary to appearances, and that we should get over it. I can say this because I have a theory to the effect that we improperly construct our philosophical theories and debates around idealized notions such that any compatibilist account of the significance of free will is renderedinadequate because it accepts that agency is given, that is, a product of forces far beyond our causal reach.

It is true, however, that all I have provided is a sketch of a theory. We cannot really say “get over it” until we have a proper non-ideal construction of the problem of free will and determinism. The foregoing suggests some desideratum on a better construction. An adequate construction of the problem of free will and determinism should: (1) be neutral with respect to which normative considerations are fundamental, (2) correctly recognize that on any plausible account of free agency, free agency will at least be partly given rather than self-created, and so (3) not rule out the compatibilist commitment to the givenness of agency.[[9]](#footnote-9) Indeed, given my earlier remarks about the givenness of agency in historical and comparative contexts, the same could be said for the construction of several relatedfree will problems.

But even without proper characterizations of (1)-(3) currently in hand, we can see how non-ideal concerns might change our priorities right now in thinking about free will and moral responsibility. It an important feature of non-ideal political theory that it can reject idealized misconstructions of a theoretical problem while *also* theorizing in terms of more local aspirational ideals (Kogelmann 2020: 243, Mills 2005: 168). When facing local threats to free agency, like oppression and manipulation, compatibilists seem well poised to be not only metaphysical realists but also metaphysical progressives who seek to theorize about how to develop meaningful agency, including but not limited to morally responsible agency, perhaps even aiming at systems that get better at creating valuable forms of agency over time or deepen the significance of our agentive contributions (cf. Barrett 2020).

To see why compatibilists should engage in this progressive metaphysical project, consider an objection from Saul Smilanksy (2013: 5-8). He argues that compatibilism is a “crazy ethics” because it implies that when we correctly punish a criminal or blame the wrongdoer, the fact that they are a wrongdoer is in some part the product of *our collective policies*, which are outside of the wrongdoing agent’s causal reach. If we had changed the antecedent causal conditions in the right way—for instance, if we alleviated socio-economic inequality—then the agent may not have committed the wrong, for they would have been a different kind of agent. What justice is this? But this objection to compatibilism is just a version of the absurdity challenge cast in second-personal moral terms. To the non-ideal compatibilist, there is nothing crazy in thinking that wrongdoers are apt targets of certain modes of treatment and that we ought to do what we can to save persons from the misfortune of becoming wrongdoers. Non-ideal compatibilists can accept that political, social, and epistemic systems partly determine the shape and nature of our free agency, and that these systems are not always, or typically, within the causal reach of individual agents. Instead of seeing this as a threat, my suggestion is that we see this as a commonsense fact and maybe even an opportunity.

Readers may be left with a lingering question. If I am right, then compatibilists can reject the view that we *should* find compatibilism absurd. But what would it look like to *not* as a matter of fact, feel that compatibilist agency is absurd? My answer to this question will reinforce the view that non-ideal compatibilists should change their priorities in thinking about free will and moral responsibility. Recall Watson’s ontological shudder and the thought that when we consider those who have suffered and so become bad people we might say, “there but for the grace of God go I.” Who am I to dismiss my grandmother’s wisdom, for her expression is not only a recognition of our fragility but is also a prayer of gratitude. Perhaps by transforming our shudders into prayers, repeated from time to time “as if it were an incantation” we can transform our metaphysical anxiety into metaphysical appreciation. (Plato, *Phaedo* 114d). The appreciation of one’s own good fortune should inspire us to seriously consider the political priorities I have gestured at above, both in making the case for compatibilism and for setting our philosophical agenda. Gratitude, I hope, will help us discern and cultivate the many significant forms of free agency the world may afford us.

1. **Conclusion**

The title of this paper is tongue-in-cheek, but the paper is a manifesto of sorts, nevertheless. I have publicly declared a set of programmatic aims, advertising a shared endeavor to other compatibilists. I have argued that compatibilism has a cost. It strikes us as unsettling. It feels wrong. It makes us seem absurd. By my lights, this has always been *the cost* of the view. I diagnosed this feeling in terms of absurdity, and then I offered a remedy. I suggested that the current free will debate is influenced by problematic idealizations that ought to be rejected. By naming and rejecting those idealizations, we can principally dispel the emotional costliness of the view. Working to dispel this costliness also provides grounds for an appealing metaphysical progressivism.

**References**

Arpaly, Nomy. 2006. *Merit, Meaning, and Human Bondage*. Princeton: Princeton University Press.

Baker, Lynne Rudder. 2003. “Why Christians Should Not Be Libertarians: An Augustinian Challenge” *Faith and Philosophy* 20 (4): 460-478.

Barrett, Jacob. 2020 ‘Social Reform in a Complex World.’ *Journal of Ethics and Social Philosophy* 17(2): 103-132.

Blackburn, Simon. 1998. ‘Relativization and Truth’ in Lewis Hahn (ed.), *The Philosophy of P.F. Strawson* (pp: 151-167). Open Court Publishing: Peru, Illinois.

Bok, Hilary. 1998. *Freedom and Responsibility.* Princeton: Princeton University Press.

Clarke, Randolph. 2005. ‘On an Argument for the Impossibility of Moral Responsibility.’ *Midwest Studies in Philosophy* 29: 13-24.

Dennett, Daniel. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge, Mass: MIT Press.

Finnigan, Bronwyn. 2022. “Karma, Moral Responsibility, and Buddhist Ethics” in Manuel Vargas and John M. Doris (eds.), *The Oxford Handbook of Moral Psychology* (7-23), Oxford University Press, Oxford.

Frankfurt, Harry. 1971. ‘Freedom of the Will and the Concept of a Person.’ *Journal of Philosophy* 68(1): 5-20.

Fischer, John Martin. 2009*. Our Stories*. New York: Oxford University Press.

Fischer, John Martin and Mark Ravizza. 1998. *Responsibility and Control: An Essay on Moral*

*Responsibility*. Cambridge: Cambridge University Press.

Gier, Nicholas F. and Paul Kjellberg. 2004. “Buddhism and Freedom of the Will: Pali and Mahayist Responses” in *Freedom and Determinism: Topics in Contemporary Philosophy*,
eds., J. K. Campbell, D. Shier, M. O’Rourke (Cambridge: MIT Press): 277-304.

Kane, Robert. 1998. *The Significance of Free Will.* Oxford: Oxford University Press.

Kogelmann, Brian. 2020. ‘The Future of Political Philosophy: Non-Ideal and West of Babel.’ *Review of Austrian Economics​* 33: 237-252.

Kristjánsson, Kristjàn. 2003. ‘Justice, Desert, and Virtue Revisited.’ *Social Theory and Practice* 29(1): 39-63.

McKenna, Michael. 2021. ‘Wimpy Retributivism and the Promise of Moral Influence Theorists.’ *The Monist* 104 (4): 510- 525.

McKenna, Michael. 2019. ‘The Free Will Debate and Basic Desert.’ *The Journal of Ethics,* 23(3): 241-255.

McKenna, Michael. 2013 “Reasons-Responsiveness, Agents, and Mechanisms”. In David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*, Vol. 1 (pp. 5-37). Oxford: Oxford University Press.

McKenna, Michael. 2008. “Ultimacy & Sweet Jane.” In Nick Trakakis and Daniel Cohen (eds\_., *Essays on Free Will and Moral Responsibility* (pp. 186-208). Cambridge: Cambridge Scholars Publishing.

Mele, Alfred. 2006. *Free Will and Luck*. New York: Oxford University

Mills, Charles. 2005. ‘‘Ideal Theory’ as Ideology.’ *Hypatia*, 20(3): 165-184.

Nagel, Thomas. 1971. ‘The Absurd.’ *Journal of Philosophy*,68(20): 716-727.

Nozick, Robert. 1981. *Philosphical Explorations*. Cambridge Mass: Harvard University Press.

Nelkin, Dana. 2013. ‘Reply to Critics.’ *Philosophical Studies*, 163: 123-131.

Nelkin, Dana. 2011. *Making Sense of Freedom and Responsibility*. Oxford: Oxford University Press.

Pereboom, Derk. 2021. *Wrongdoing and the Moral Emotions*. Oxford: Oxford University Press.

Pereboom, Derk. 2019. ‘What Makes the Free Will Debate Substantive?’ *The Journal of Ethics,* 23(3): 257–264

Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.

Plato. 1997. *Plato: Complete Works*. John M. Cooper ed. Indianapolis, IN: Hackett Publishing.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass: Harvard University Press

Rogers, Katherine A. 2004. “Augustine’s Compatibilism.” *Religious Studies* 40 (4): 415-435.

Russell, Paul. 2022. ‘Moral Responsibility and Existential Attitudes’ in Dana K. Nelkin and Derk Pereboom (eds.), *The Oxford Handbook of Moral Responsibility* (pp. 520-541). Oxford: Oxford University Press.

Russell, Paul. 2017. *The Limits of Free Will*. Oxford: Oxford University Press

Sartorio, Carolina. 2016. *Causation and Free Will*. Oxford: Oxford University Press.

Sartorio, Carolina. 2014. “The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will”. In Alfred Mele (ed.), *Surrounding Free Will: Philosophy, Psychology, Neuroscience* (255-273). Oxford University Press: Oxford

Scanlon T.M. 2008. *Moral Dimensions: Meaning, Permissibility, Blame*. Cambridge, Mass.: Harvard University Press.

Shoemaker, David. 2018. “You Oughta Know: Defending Angry Blame.” In Myisha Cherry and Owen Flanagan (eds)., *The Moral Psychology of Anger* (pp. 67-88). New York: Rowman and Littlefield.

Shoemaker, David. 2015. *Responsibility from the Margins*. Oxford, UK: Oxford University Press.

Sen, Amartya Kumar. 2009. *The Idea of Justice*. Cambridge Mass: Harvard University Press.

Sider, Theodore. 2009. “Ontological Realism.” In David Manley, Ryan Wasserman, and David Chalmers (eds.), *Metametaphysics: New Essays on the Foundations of Ontology* (pp. 384-423)*.* Oxford: Oxford University Press.

Sideret, Mark. 2017. “Buddhist Paleocompatibilism.” In Rick Repett (ed.), *Buddhist Perspectives on Free Will: Agentless Agency?* (pp. 33-147) London: Routledge.

Smilanksy, Saul. 2013. “Free Will as a Case of “Crazy Ethics”. In Gregg Caruso (ed.), *Exploring the Illusion of Free Will and Moral Responsibility* (pp. 103-120). Maryland: Lexington Books.

Smilanksy, Saul. 2003. ‘Compatibilism: The Argument from Shallowness.’ *Philosophical Studies*,115: 257-82.

Srinivasan, Amia. ‘The Aptness of Anger.’ *The Journal of Political Philosophy* 26 (2):123-144.

Strawson, Galen 1994. ‘The Impossibility of Moral Responsibility.’ *Philosophical Studies* 75:5-24.

Strawson, P.F. 1962. ‘Freedom and Resentment.’ *Proceedings of the British Academy* 48: 187-211.

Stump, Eleonore. 2014. “Augustine on Free Will” in Eds. Eleonore Stump and David Vincent Meconi, *The Cambridge Companion to Augustine*, 2nd edition (pp. 166-186). Cambridge Unviersity Press: Cambridge.

Stump, Eleonore. 2003. *Aquinas*. New York, NY: Routledge.

Valentini, Laura. 2012. ‘Ideal vs. Non‐Ideal Theory: A Conceptual Map.’ *Philosophy Compass* 7(9): 654-664.

Vargas, Manuel. 2013. *Building Better Beings*. Oxford: Oxford University Press.

Vihvelin, K. 2013. *Causes, Laws, & Free Will: Why Determinism Doesn’t Matter*. New York: Oxford University Press.

Wallace, Robert H. 2021. “The Tension in Critical Compatibilism”. *Ethical Theory and Moral Practice* 24(1): 321-332.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA.: Harvard University Press.

Watson, Gary. 1987/2008. ‘Responsibility and the Limits of Evil: Variations on a Strawsonian Theme’ in McKenna, Michael and Paul Russell (eds.), *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment"* (pp. 19-36). Ashgate, Burlington, VT: Farnham, England.

Watson, Gary. 1975. “Free Agency.” *Journal of Philosophy* 72(8): 205-220.

Wolf, Susan. 2015. *The Variety of Values: Essays on Morality, Meaning, and Love*. Oxford: Oxford University Press.

1. My sincere thanks to Robert J. Hartman, Michael McKenna, Jeremy Reid, David Shoemaker, Jacob Sparks, Hannah Tierney, Ava Wright, two anonymous referees, and the participants of the seventh New Orleans Workshop on Agency and Responsibility for help on this paper. [↑](#footnote-ref-1)
2. For example, see: Mele (2006: 17), McKenna (2008:187), Pereboom (2014: 1-4) and Sartorio (2016: 7). [↑](#footnote-ref-2)
3. This is not how Nagelthinks about the problem of free will. More on this in section 4. [↑](#footnote-ref-3)
4. An alternative but equivalent formulation of the absurdity challenge can be expressed in terms of moral luck, since accepting the givenness of agency is equivalent to accepting a kind of causal luck. [↑](#footnote-ref-4)
5. This is so even for versions of compatibilism that ground free will in the fact that an agent is able to do otherwise. If this ability claim is compatible with determinism, then *how* an agent exercises their ability to do otherwise will be explicable in terms of causal antecedents. If the causal antecedents were different, then the agent would have exercised their ability in a different way. So, it had better be the case that the causal antecedents were of the right sort. [↑](#footnote-ref-5)
6. My thanks to a referee for raising this point. [↑](#footnote-ref-6)
7. Thanks to David Shoemaker for this turn of phrase. [↑](#footnote-ref-7)
8. E.g., Daniel Dennett (1984), R. Jay Wallace (1994), T.M. Scanlon (2008), Hillary Bok (1998) Manuel Vargas (2013), and Paul Russell (2017). [↑](#footnote-ref-8)
9. Despite my disagreement with howVargas constructs the problem, and so with his recommendations (in 2013, *inter alia*), his approach meets these three criteria. See (2013: 128) for criteria (1) and (2013: 244-24) for criteria (2). Criteria (3) is met throughout. [↑](#footnote-ref-9)