Received: 23 July 2020 | Revised: 22 February 2021 | Accepted: 10 March 2021

DOI: 10.1111/phib.12227

ORIGINAL ARTICLE

ANALYTIC PHILOSOPHY WILEY

Compatibilism from the inside out[1]

Andrew M. Bailey

Yale-NUS College, Singapore

Correspondence
Andrew M. Bailey
Email: andrew.bailey@yale-nus.edu.sg

Funding information
Yale-NUS College, Grant/Award Number: R-607-000-305-115

Abstract

In this article, I focus on internal dimensions of moral responsibility. I argue that if there are such dimensions, then moral responsibility is compatible with determinism.

KEYWORDS

compatibility, determinism, free will, moral responsibility

1 | INTRODUCTION

If you are like me, you are blameworthy for some goings-on in our world—broken promises, hurt feelings, frustrated preferences, missed deadlines, and worse things besides. And if you are a little less like me, you are praiseworthy for some goings-on too—happy children, well-tended gardens, delicious meals, and so on. We are also, I venture to guess, praiseworthy, or blameworthy for some goings-on *inside our own heads*—desires selfish or kind, hatred or love, good or bad intentions, resolutions to act, willings, and so on. We have done or thought or felt good and bad things—*in the mind alone*, if you like—and for such *internal* affairs we are eligible for praise or blame regardless of their worldly consequences.

In this article, I focus on these *internal* dimensions of moral responsibility. I will argue that if such dimensions are real, then moral responsibility is compatible with determinism. Those inclined to affirm the antecedent with me will find in this article a new argument for compatibilism about moral responsibility and determinism.

2 | THE INSIDE OUT ARGUMENT

Let us say someone is *morally responsible* just if she is an apt candidate for praise or blame. *Determinism* is true just if that there is, at any time, exactly one physically possible future. Moral responsibility and determinism and *compatible* just if possibly, someone is morally responsible and determinism is true.

Here is a rough statement of my *Inside Out* argument for the conclusion that moral responsibility is compatible with determinism (a more careful formulation will follow):

There are at least *some* items—facts, objects, properties, states of affairs, events, take your pick—that are both internal to us and that are objects of moral responsibility. Internal: they live and move and have their being within our boundaries. In the parlance of our times, the properties we enjoy when and because they're there are *intrinsic* to us. So anything exactly alike us in all intrinsic respects shares these features too. Objects of moral responsibility: they are items for which we are apt candidates for praise or blame.

Consider an unusually *unkind intention* that one Jo harbors in her heart.

Might such an intention be something for which Jo is an apt candidate for blame? It seems so; we often show great perfervidity in blaming people for bad intentions, even ones that do not lead to action. If you do not think bad intentions are appropriate objects for such blame, feel free to replace my talk of intentions in the sequel with some other suitable mental item for which one might be aptly blamed, such as an unkind state or quality of will or a decision or resolution to act.[2]

Might her bad intention also be intrinsic to Jo? Again, it seems so; Jo's intrinsic duplicates will vary across many dimensions, but all of them will harbor some particularly unkind intention and will therefore be equally apt candidates for blame. To put all this a little differently: if you wanted to change the fact that Jo is blameworthy for her unkind intention, you would need to change something about Jo herself, by eliminating that intention altogether, somehow shifting it around in her mind, or even removing Jo from the picture entirely. Merely dropping Jo into different surroundings would not do the trick.

Determinism, moreover, is not *intrinsic* to us. More carefully, *being such that determinism is true* is not intrinsic to us. If determinism is true that would be at least partly a matter of the *general* structure of space, time, and the laws of nature—and not just a matter of what's going on inside you, or me, or any one of us on her own. Likewise, if determinism is false. This much seems plain. An extension seems plausible too: we (or intrinsic duplicates of us) could exist under either determinism or its denial. To see why this extension is plausible, suppose our world enjoys or suffers under "99.9% deterministic" laws of nature. Such laws do not *guarantee* a unique future given any state of the world at a time, but they do make it *exceedingly* likely that a particular future will in fact unfold. Could something internally just like you have existed, had the laws been "100% deterministic" instead? It seems so. Even though small variations in the laws might make for very big differences in some respect or other (as incompatibilists maintain), they *need not* make for big differences in what is going on *inside* your head and mine.

The theses expressed so far imply compatibilism about moral responsibility and determinism. Where *Jo* is one of us, and *the bad intention* is some bad intention she holds, here is one way to regiment the Inside Out argument:

1. *Being morally responsible for the bad intention* is intrinsic to Jo.
2. *Whether determinism is true* is extrinsic to Jo (i.e., either *being such that determinism is true* is extrinsic to Jo or *being such that determinism is false* is extrinsic to Jo).
3. If *whether determinism is true* is extrinsic to Jo, then Jo has an intrinsic duplicate at a deterministic world.
4. Therefore, Jo has an intrinsic duplicate at a deterministic world (from 2 to 3).
5. Therefore, Jo has a morally responsible duplicate at a deterministic world (from 1 to 4).
6. Therefore, possibly: someone is morally responsible and determinism is true (from 5).

My case for the premises is implicit in the intuitive summary of the argument given above, but I will now make things more explicit and add some extra supporting reasoning for lines one through three; the rest follow by valid deduction.

2.1 | Premise one: intrinsic responsibility

Reflect on what might vary given variation in Jo's surrounding circumstances. Shift things around a bit, and the bad intention might have better results in the world; or worse; or no external results at all. But shift as you may, you will not change the bad intention itself. Nor will you change the appropriate stance toward Jo and her bad intention. This is the hallmark of an intrinsic property—it is impervious to external variation. So it is plausible to suppose that Jo's moral responsibility for the bad intention is intrinsic to her. I think this kind of thought experiment supports premise one. But I concede that I have no knock-down *argument* for the premise. That's why it is a premise and not a conclusion. I have no decisive proof. What I *can* do, however, is point out that the premise has already been accepted by some avowed incompatibilists[3] and defend the premise against objections. More on that below.

Many of the items for which we are paradigmatically morally responsible—broken hearts, unjust deaths, cultivated gardens, virtuous children, saved lives, and so on—are *external* to us.[4] You cannot break a promise to someone else, for example, unless someone else is there; but *being such that someone else is there* is not intrinsic to you. Is the "moral responsibility internalism" at play here, then, implausible? Think about moral luck, and you will feel the force of this question. If moral luck—especially "resultant" moral luck—is possible, then variation in external circumstance alone can make for dramatic variation in facts about who is morally responsible for what. I concede these points.[5]

But notice why they do not cut against my argument. There are many plausible cases of extrinsic moral responsibility, like responsibility for breaking a promise—and if moral luck is possible, there are even more cases besides. But I maintain that there are *some* items for which we are responsible that are intrinsic to us. Some disciples of Abelard and Kant seem to think that *all* objects of ultimate moral responsibility are internal to us (mental acts or states of will, for example); but the operative assumption here—that there are *some*—is comparatively unassuming and plausible. The conclusion of the argument is modest too. It does not claim that moral responsibility for *anything at all* is compatible with determinism, just that *someone* could be morally responsible for *something*—even if determinism is true.

I have emphasized that variation of external circumstances does not seem to make a difference to Jo and her bad intention, though it may make a great deal of difference to other worldly matters. This strategy is consonant, I note, with familiar "Frankfurt-style cases," where variation of the presence of a counterfactual intervener—an extrinsic factor—does not seem to make a difference to moral responsibility facts, though it may make a great deal of difference to facts about what the hapless subjects of those cases can do.[6]

2.2 | Premise two: extrinsic determinism

Whether determinism is true is in part a matter of whether there are laws of nature, and whether they are deterministic or not. I will now argue that premise two may be further bolstered by considering particular accounts of what laws of nature *are*. The considerations below offer independent reasons to affirm the premise.

Laws of nature are said by some—Humeans—to consist in generalizations about and to depend on all the various local matters of fact.[7] Whether it is a law that such-and-such depends on what's going on *all across the cosmos*. Laws thus enjoy a large base, extrinsic to any one item in the legion of concreta that are. So on this account, properties encoding those laws (*being such that it is a law*

that such-and-such, say) are not intrinsic to any particular individual; they are not intrinsic to Jo. Such properties need not even be intrinsic to the cosmos as a whole, for a cosmos exemplifying that property might be duplicated without preserving the property; to get this result, just add in some "extra" reality with counter-instances to *such-and-such*.[8]

Laws of nature are said by others—non-Humeans—to consist in something rather different from the vast mosaic of local matters of fact. Leading candidates include relations between universals or unusually potent generalizations that somehow *govern* local matters of fact.[9] On these accounts, too, properties encoding those laws will be extrinsic to items like Jo. For whether Jo has one of those properties will depend on the universals and their relations or those governing generalization—relations and generalizations quite external to Jo.

Some non-Humean accounts of laws will undercut my case for premise two. Here is one:

> Laws are relations between universals. And what a universal *is* consists partly in how it figures into the laws (and the laws into which it figures). Variation in the laws—for example, from indeterminism to determinism—guarantees variation in what universals are instanced. So: if Jo lives under the governance of indeterministic laws, she in fact has no proper duplicate at a deterministic world (and vice versa). That duplicate might *look very much like her*, so to speak, but it will in fact instance distinct universals. All of Jo's features are essentially embedded in a network of indeterministic laws and cannot be instanced in a world with deterministic laws. Here's the payoff. On the package of views being considered here, it is not obvious that, if indeterminism is true, then *being such that indeterminism is true* is extrinsic to Jo. Rather, indeterminism is "built into" Jo because the intrinsic features she enjoys (universals) can be instanced only if indeterminism is true. Jo *could not* have a duplicate at a deterministic world after all.[10]

An intriguing objection. I offer three replies.

First, an observation. Note that only on very *specific* packages of views do we have an undercutting objection to premise two. For example: (a) There are universals; (b) A non-Humean view about laws is correct; (c) Laws are relations between universals; (d) What a universal *is* consists partly in how it figures into the laws (and the laws into which it figures); (e) Whether a law is deterministic or not is essential to that law.

No one of these will alone do the trick; my objector needs the Whole Package. If it turned out that incompatibilists had to—to diagnose the error in the Inside Out Argument, I mean—endorse not just one, but *all* of these theses to defend their view that would be a surprising and noteworthy fact.

Second, there seem to be good reasons to doubt the Whole Package. The Whole Package implies that *every* contingent characterization is *maximally* modally fragile—when it comes to determinism and its denial. On the Whole Package, *any* contingent feature instanced in a deterministic world can *only* be instanced at deterministic worlds; and so also for features instanced in indeterministic world. I grant that some degree of modal fragility seems right. Change the laws a lot and you will surely change what universals are instanced. But must *any* variation in the laws guarantee variation in what universals are instanced? This is less plausible. Suppose it is a law that *F is 99.9999% likely to be followed by G* and that there are some Fs. Must we insist that the laws *could not* have instead dictated that *F is 100% likely to be followed by G*, in a world with some Fs? I do not think so. And so I am unconvinced of the Whole Package.

2.3 | Premise three: Duplication

Premise three is, I shall suppose, true by definition—or something like that. It is an instance of a general and axiomatic relation between intrinsicality, extrinsicality, and duplication: if a (contingent) property is extrinsic to an item, then there are worlds containing a duplicate of that item without the property.[11] I am here assuming that *being such that determinism is true* or *being such that determinism is false* would only contingently characterize Jo—determinism is a thesis that is both possibly true and possibly false, and Jo could exist in either case. Necessitarians—who maintain that, of necessity, all truths are necessary truths—will not accept this assumption. But that is not quite relevant to the Inside Out argument. For necessitarians who accept the possibility of moral responsibility are already committed to compatibilism; if *x* is a necessary truth and *y* is possibly true, then the conjunction of *x* and *y* is also possibly true.

For ease of initial presentation, premises two and three deploy "whether determinism is true"—not exactly a canonical and transparent name for a property (unlike, say, "being such that determinism is true"). To see why this obscurity is not damning, notice two cases and what follows. Either determinism is true, or it is not; that is, Jo either has the property *being such that determinism is true*, or has the property *being such that determinism is false*. If she has that former property, then we already have a case of someone who is morally responsible despite determinism's being true. If she has the latter property, then it is her duplicate (at a deterministic world) who supplies us with the needed case of someone who is morally responsible despite determinism's being true. Either way, the conclusions follow.

3 | OBJECTIONS

I have already recommended the premises. Let me now clarify and defend them. My goal is to show that the Inside Out argument—even if not decisive—is still plausible, lends support to its conclusion, and raises the price of the theory that determinism is *in*compatible with moral responsibility.[12]

Objection 1: We are much more modally fragile than the argument presupposes. For were the laws different in any respect, we would not exist at all. The same applies to Jo; jigger with the laws, and she would cease to be. Or, more modestly, we cannot be confident that Jo could exist under different laws, and so should not affirm a premise requiring as much.

Reply: My premises do not require that *Jo herself* could exist under both determinism and its denial, just that she has intrinsic duplicates that are impervious to such alterations in the laws. For the objection to succeed, it must say that no *duplicate* of Jo could exist at all under different laws, and this form of essentialism is not nearly so compelling.

Objection 2: The third premise appears to rely on a principle of recombination, according to which distinct patches of reality may be freely recombined. So, the thought goes, since the patches of reality that fix whether determinism is true or not are distinct from Jo's own patch, there are worlds with Jo's patch intact in which the determinism-fixing patches vary. But this principle of recombination is false. Here's one counterexample among many: An all-powerful, all-loving, and all-knowing deity is incompatible with sufficiently and gratuitously evil patches, even though such evil patches are distinct from any deity's patch.

Reply: A Ludovician principle of recombination may indeed support premise three. But that is not why *I* recommend premise three. I do not appeal to any wholly general principles here. I affirm premise three, instead, because—though "turning determinism on or off" *could* make for rather big differences in a world—it *need not* make for *internal* ones.

Objection 3: The first premise presupposes that someone is morally responsible for something. But this is dubious and at least dialectically infelicitous, given the growing prevalence of skepticism about moral responsibility.

Reply: Those who are convinced that moral responsibility is *impossible* may find little to like in my argument. If you are one of those skeptics, I recommend that you take my argument as advancing this interesting conditional with a false antecedent: *if* moral responsibility is possible, then it is compatible with determinism.

More modest forms of skepticism (according to which, for example, moral responsibility is *rare* or merely *contingently* absent) are compatible with the substance of the argument. For example, instead of looking for *actual* cases of responsibility, we could simply suppose that they are *possible*, and run the argument accordingly.

Objection 4: The Inside Out argument presupposes a *snapshot* theory of moral responsibility, according to which the factors by virtue of which someone is morally responsible could all be captured in a momentary snapshot of her intrinsic states. Such a theory is, at any rate, the most promising way of motivating premise one. But incompatibilists (and others) have long rejected snapshot theories, and for good reason. It is not just your internal states that make you morally responsible; it also matters *how you got into those states*—whether by manipulation or free choice, for example.

Reply: premise one does not require that *being morally responsible for the bad intention* is intrinsic to Jo at exactly one moment; nor should it be read as though it were indexed to a time. For a property can be intrinsic to something even if not *temporally intrinsic* to the thing at some moment or other. That Spot the Spaniel is *spotted at t and not-spotted at t**, for example, is intrinsic to Spot, even though it is neither temporally intrinsic to Spot at *t* nor at *t**. Think of *being morally responsible for the bad intention*, then along those lines: it is intrinsic to Jo (or to her *total history*, if you like), even if it requires that she be in various states at various distinct times.

Objection 4 may be remixed. Perhaps, the problem is that moral responsibility has a historical condition that requires environmental cooperation. And any condition that requires environmental cooperation is not intrinsic. That moral responsibility involves such historical or environmental conditions may even be a consequence of a general theory.

Reply: that compelling theories of moral responsibility are incompatible with premise one is some evidence against it. But the premise may retain its plausibility. Consider someone who is a paradigm of moral responsibility for some mental item, like a bad intention. Would every intrinsic duplicate of that paradigm also be morally responsible, even if in a different environment? I think so. And I think reflection here supports premise one, even if various recondite theories of moral responsibility tell against it. We need not reason only from the top down—from theory to a judgement about cases. We may, instead, reason from the bottom up—from a judgement about cases to some more abstract theory.

Objection 5: The Inside Out argument slides all-too-quickly from "the bad intention is intrinsic" to "Jo's blameworthiness for the bad intention is intrinsic." Why should the former entail the latter? There might, after all, be conditions on moral responsibility that are extrinsic; an epistemic condition on moral responsibility, for example, might require knowledge of something or other—a paradigmatically extrinsic state.

Reply: The slide is warranted, at least in the present case. Suppose Jo is indeed responsible for the bad intention, and that the bad intention is intrinsic to Jo. How could we remove Jo's responsibility? It seems to me that we would have to re-arrange, somehow, Jo's internal makeup, whether by removing the bad intention or radically restructuring its place within her internal economy. On extrinsic epistemic conditions on moral responsibility: if there are such requirements, they most plausibly hold with respect to *external* objects of moral responsibility, like the worldly consequences of one's choices. But

external objects of moral responsibility are not at issue here, and so requirements on such are not at issue either.

Objection 6: The Inside Out argument, roughly, has an *internal* move (slogan: "moral responsibility is intrinsic") and an *external* move (slogan: "whether determinism is true is extrinsic"). But these moves cut against each other. Once you have given the incompatibilist good reason to accept one move, you will have given her equally good reason to reject the other. Any incompatibilist who agrees with you, for example, that moral responsibility is intrinsic must disagree with you about the extrinsicality of whether determinism is true.

Reply: In a way, I agree with the objection. For it points out an undeniable feature of valid arguments. When some premises p and q imply a conclusion c, there will always be a valid argument proceeding from not-c and p to not-q. But despite the fact that valid arguments can always be *flipped* in this way, they still have some uses. I will note three.

First, they expose the implications of accepting various premises or denying various conclusions. Even when they do not *convince*, they still *clarify* various logical connections. The Inside Out argument points to a heretofore hidden cost of incompatibilism about moral responsibility and determinism: denying one or more plausible—or at least, reasonable—premises about which features are intrinsic.

Second, they can persuade *neutral* audiences to accept the conclusions. Even if no *incompatibilist* will accept my premises, the case I have made for them shows that they—and the conclusions they entail—can be reasonably accepted by those not already committed to incompatibilism.

Third, they can give audiences, neutral or otherwise, reason to *increase confidence* in their conclusions, thereby coming closer to, if still a bit short of, full confidence, belief, or acceptance. You might be such a reader. If so, you should lean more toward compatibilism now than you did before reading this article.

I, for one, find the premises plausible enough, and so I accept the conclusion that someone could be both morally responsible and determined.

4 | THE PROSPECTS FOR COMPATIBILISM

Wise compatibilists will do two things. First, they will give a diagnosis of where, exactly, incompatibilist arguments go wrong. Second, they will give positive account of how, exactly, compatibilism could ever be true. This article does not help much when it comes to the first task. But it does offer helpful pointers—and some cautionary notes—for philosophers interested in the second.

I imagine a compatibilist theory of moral responsibility beginning inside and working outwards. If we could be morally responsible for at least some of the goings-on in our heads—as the argument of this paper suggests, and even if determinism is true—then perhaps such are the *basic* cases of moral responsibility. Non-basic cases (for example, moral responsibility for the *consequences* of one's intentions, which consequences roam free of our heads) may impose additional metaphysical requirements (for example, suitable and foreseeable connection with basic cases of moral responsibility, as when an intention bears external fruit in just the right way). This, in turns, raises interesting questions about what that suitable connection must be, and whether *it* is compatible with determinism.

Here is a suspicion. Basic cases of moral responsibility are indeed compatible with determinism; even if determinism is true, we could be morally responsible for goings-on in our heads (or minds, if you prefer). Compatibilism about moral responsibility and determinism is therefore true. But *non*-basic cases are more metaphysically demanding, and their demands are not so obviously satisfied. It remains an open question whether *they* could ever come to be—whether in worlds where the past and

laws conspire to fix our fate, or in worlds where they do not and where we are left to own devices. The position defended in this article is thus amenable to various flavors of skepticism about moral responsibility, including even the conviction that no one is ever blameworthy or praiseworthy for any worldly consequences of her internal states.

5 | CONSEQUENCES AND CONNECTIONS

Suppose the Inside Out argument is sound. What then? Does it matter? And how does all of this bear on those old chestnuts about liberty and necessity—or free will and determinism? I will close with tentative answers to these questions.

Here's one reason to care about the fate of the Inside Out argument. If you are like me, you do not just praise and blame people. You are probably—at least sometimes—rather enthusiastic about all that. We *like* to judge each other. It *feels good* to toss a little disapprobation in the direction of evil-doers, to give a raucous cheer when someone does right. The incompatibilist about moral responsibility and determinism thinks we can reasonably do this *only if determinism turns out to be false*. After all, on her theory, if determinism is true, then no one is responsible for anything. The incompatibilist thus maintains that moral responsibility depends on a particular physical or metaphysical hypothesis. If my argument is correct, our status as moral responsible creatures (and by extension, our practices of praising, blaming, and so on) is not in so precarious a position.[13] Even when Athena herself whispers to you that determinism is true, you need not concede that no one is responsible for anything, nor need you stop praising and blaming. I will not say this is *good* news without qualification. But for those inclined toward praise and blame, the Inside Out argument brings some comfort.

The comfort needs qualification. I have distinguished basic (internal) and non-basic (external) cases of responsibility. I have also suggested that the Inside Out argument establishes only that the basic cases are compatible with determinism. We could be apt candidates for praise or blame *in those basic and internal cases*—even given determinism. Our moral responsibility in those cases is impervious to determinism. It does *not* follow that could be apt candidates for praise or blame *in non-basic and external cases*, given determinism. Our moral responsibility in non-basic cases *need not* be impervious to determinism. So if determinism is a live hypothesis—one we might well discover in the course of physical, metaphysical, or theological inquiry—then my qualified comfort is this: go ahead and keep on praising and blaming people for the goings-on in their heads. You need not worry then about whether determinism is true. But be much more reserved about praise and blame for the worldly consequences of those goings-on; and, absent further argument, be prepared to give up on blaming or praising in those non-basic cases should it ever turn out that determinism is true.

For all the Inside Out argument teaches us, of course, there may be *other reasons*—independent of determinism, I mean—to be skeptical about moral responsibility. Here, then, is another reason to care about that argument. If it is sound, then those interested in evaluating praise and blame have reason to investigate these *other reasons*. There are purely philosophical questions here. For example: what relations between intentions, beliefs, desires, resolutions to act, willings, and other mental states make for internal and basic cases of moral responsibility? And what could possibly extend our responsibility out there into the external world in non-basic cases? There are empirical or scientific questions as well. For example: do those relations *in fact* obtain? How about the factors that might extend our responsibility out into the world—do those factors *in fact* obtain? The Inside Out argument, then, points us in interesting and productive directions for future research.

The Inside Out argument offers no *direct* and *conclusive* answers to venerable (or ancient, at any rate) questions about free will and determinism. But it does have some implications here as well.

I am here neutral on the question of how to best understand free will, and in particular, whether free will is *being able, in a given situation, to exercise the logically strongest kind of control necessary for moral responsibility* or whether it is, instead, *being able, in a given situation, to do something or to not do it*. I note that on the former account—but not on the latter—having free will is plainly necessary for being morally responsible.

If moral responsibility *requires* free will—if, necessarily, someone is morally responsible for something only if she, at some time or other, has free will—then the Inside Out argument shows that free will is indeed compatible with determinism. This consequence follows by elementary modal reasoning: if x entails y and the conjunction of x and z is possibly true, then the conjunction of y and z is also possibly true). So given this consequence, those of us who care a great deal about whether we have free will can adopt the irenic stance explained above. We need not worry about some dramatic discovery in physics, metaphysics, or theology—that determinism is true, I mean—and its possible implications for our freedom.

If free will is *not* required for moral responsibility, then I think the Inside Out argument suggests something a little more peculiar. Perhaps, we can rest a little easy when it comes to moral responsibility. But we cannot rest so easy when it comes to free will. Should this bother us? It should, I submit, *only if* free will is required for or makes possible good things wholly disjoint from moral responsibility. Possible candidates for this office include genuine love, real relationships, and meaning in life. The Inside Out argument, then, recommends that philosophers continue thinking about free will, not just in relation to determinism or moral responsibility, but also with an eye toward other goods it might enable.

ORCID


Andrew M. Bailey https://orcid.org/0000-0002-6030-4896

FOOTNOTES

[1] I have subjected many to nascent versions of the arguments defended here. For advice and trenchant criticism, I thank anonymous referees, Alex Arnold, Nathan Ballantyne, Justin Capes, Justin Coates, EJ Coffman, John Martin Fischer, John Hawthorne, Garrett Pendergraft, Alex Pruss, Josh Rasmussen, Mike Rea, Brad Rettler, Michael Robinson, Amy Seymour, Alex Skiles, Philip Swenson, Patrick Todd, Neal Tognazzini, Peter van Elswyk, Fritz Warfield, Dean Zimmerman, and an audience at Yale-NUS College. Astute readers might guess that many of the ideas in this article are inspired by John Martin Fischer and the semi-compatibilist program he's advanced over the last few decades. Such guesses would be correct.

[2] For helpful discussion of the kind of mental items for which we can be morally responsible, see Smith (2005) and (2008).

[3] See, for example, Pruss (2019).

[4] For arguments that everything relevant to moral responsibility is internal us and that we are morally responsible *only* for inner willings, see Khoury (2018).

[5] Premise one is consistent, then, with the arguments in Ciurria (2015). Those arguments establish at most that many (but not, of necessity, *all*) cases of moral responsibility involve the external world in various ways. My premise is similarly consonant with, but does not require, strongly internalistic views about moral responsibility like that in Zimmerman (2016) and (2017).

[6] Frankfurt (1969).

[7] Lewis (1986).

[8] Humeans about laws may have special reason to resist my first premise. For on those views, Jo's *abilities* may well depend on all sorts of extrinsic matters of fact (e.g., whether her environment

cooperates with her intentions in the right way), and those abilities may partially determine whether Jo is indeed responsible for the bad intention. I do not think this is a serious barrier to the success of my main project here for two reasons. First, the Humean *need not* insist on any straightforward relationship between responsibility and ability; she could, for example, deny any entailment at all. Second, and more importantly, Humean views of laws support other, independent arguments for compatibilism about both moral responsibility and determinism and free will and determinism; on those, see Beebee and Mele (2002.

[9] Armstrong (2016), Dretske (1977), and Tooley (1977).

[10] Thanks to an anonymous referee for suggesting this objection.

[11] This is not to say that the notion of intrinsicality is without its own problems; see Marshall (2016) for just a few of them.

[12] Thanks to various anonymous referees for supplying all of these objections.

[13] The argument of this section bears obvious resemblance (and owes something to) the view that incompatibilism holds our agency hostage or suggests that it hangs by a thread. See, e.g., Fischer (1999): 129.

REFERENCES

Armstrong, D. M. (2016). *What is a Law of Nature?*. Cambridge University Press.

Beebee, H. & Mele, A. R. (2002). Humean compatibilism. *Mind*, *111*(442), 201–223.

Ciurria, M. (2015). Moral responsibility ain't just in the head. *Journal of the American Philosophical Association*, *1*(4), 601–616.

Dretske, F. I. (1977). Laws of nature. *Philosophy of Science*, *44*(2), 248–268.

Fischer, J. M. (1999). Recent work on moral responsibility. *Ethics*, *110*(1), 93–139.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, *66*(23), 829.

Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell.

Marshall, D. (2016). The varieties of intrinsicality. *Philosophy and Phenomenological Research*, *92*(2), 237–263.

Pruss, A.. (2019). Internalism About Non-Derivative Responsibility. Blog post. URL: http://alexanderpruss.blogspot.com/2019/03/internalism-about-non-derivative.html

Smith, A. M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, *115*(2), 236–271.

Smith, A. M. (2008). Control, responsibility, and moral assessment. *Philosophical Studies*, *138*(3), 367–392.

Tooley, M. (1977). The nature of laws. *Canadian Journal of Philosophy*, *7*(4), 667–698.

Zimmerman, M. J. (2016). Moral responsibility and the moral community: Is moral responsibility essentially interpersonal? *The Journal of Ethics*, *20*(1–3), 247–263. https://doi.org/10.1007/s10892-016-9233-x

Zimmerman, M. J. (2017). Strawson or Straw Man? More on Moral Responsibility and the Moral Community. *The Journal of Ethics*, *21*(3), 251–262.

How to cite this article: Bailey AM. Compatibilism from the inside out[1]. *Analytic Philos*. 2021;00:1–10. https://doi.org/10.1111/phib.12227