Journal of the American Philosophical Association (2025) 1–19
DOI:10.1017/apa.2025.10016

No Easy Compatibilism

ABSTRACT: *Traditional compatibilists respond to the Consequence Argument by denying either the fixity of the past or by denying the fixity of the laws, neither of which is without theoretical cost. Recently, however, several authors—Christian List (2019b), Scott Sehon (2016), and Ned Markosian (2012)—have introduced novel approaches to free will that, they claim, imply that determinism is no threat to free will and, thus, that free will and determinism turn out to be compatible. The strategies employed by these authors differ considerably, with one influenced by psychology and decision theory, another by traditional philosophy of action, and the other by the metaphysics of causation. Nevertheless, this article aims to show, first, that these approaches share a common thread and, second, that each implicitly requires us to give up either the fixity of the past or the fixity of the laws (but without making this explicit or explaining how their views are preferable to traditional compatibilist views with respect to giving up the fixity of the past/laws). The article concludes with an important lesson that may be learned from our survey of these novel and interesting, even if ultimately unsuccessful, approaches to compatibilism.*

KEYWORDS: compatibilism, free will, Christian List, Ned Markosian, Scott Sehon

1. Introduction

One of the main debates (perhaps *the* main debate) in the philosophical literature concerning free will is the debate about whether or not free will is compatible with causal determinism. Many philosophers conceive of free will as requiring *leeway*, i.e., the freedom (or ability) to do otherwise than what one actually does. If determinism is true, then a proposition describing the intrinsic state of the world in the distant past in conjunction with the laws of nature entails a proposition describing the intrinsic state of the world at any subsequent time. So, if determinism is true and you are reading this sentence now, then the proposition describing the entire state of the world now (including your reading this sentence) was entailed by the laws of nature and a proposition describing the state of the world a million years ago. It would seem, then, that in order for us to have free will in a deterministic world, we would need to be free to act in a way that is inconsistent with the actual past, or we would need to be free to act in a way that is inconsistent with the actual laws. But it is widely believed that the past is fixed and thus that we are not free to act in a way that is inconsistent with the actual past. Moreover, it is widely believed that the laws of nature are fixed and thus that we are not free to act in a way that is inconsistent with the actual laws. So, if determinism is true, then we lack free will (understood as requiring the freedom to do otherwise). This argument has been

dubbed the "Consequence Argument" by Peter van Inwagen (1983), and it has been developed and discussed by Ginet (1990) and Fischer (1994), among many others.

Philosophers have been divided on how best to respond to this sort of argument. On the one hand, many accept the soundness of the argument, some (libertarians) maintaining that we have free will and others (free will skeptics) denying that we have free will. There are also those (source compatibilists) who accept (or at least can accept) the soundness of the argument but who maintain that moral responsibility, and whatever freedom it requires, is nevertheless compatible with determinism. On the other hand, traditional compatibilists reject the soundness of the argument, maintaining that, for all we know, we might have the freedom to do otherwise even if determinism is true, and they do this either by denying the fixity of the past or by denying the fixity of the laws. But, given the plausibility of the fixity of the past and the fixity of the laws, these compatibilist responses come with a theoretical cost that strikes many philosophers as too high a price.

Recently, however, several authors have introduced novel approaches to free will that, they claim, imply that determinism is no threat to free will and, thus, that free will and determinism turn out to be compatible. Christian List (2019b) argues that determinism at the physical level is consistent with the indeterminacy needed for free will at the agential level, and so free will is compatible with determinism. Scott Sehon (2016) argues that, if a non-causal (or teleological) account of action explanation is correct (as opposed to the widely held causal account), then determinism (which is about *causes*) is no threat to free will. And, finally, Ned Markosian (2012) argues that appealing to the fundamental role of agents as causes ("agent-causation") provides the compatibilist with a powerful reply to the argument for incompatibilism, a reply that is unavailable on the typical event-causal model assumed by other compatibilists. While the strategies employed by these authors differ considerably (one influenced by psychology and decision theory, another by traditional philosophy of action, and the other by the metaphysics of causation), these authors share the aim of finding a simple escape from determinism's threat to free will by attending to some previously neglected factor.

We call these routes to compatibilism "easy" because they seek to avoid the dialectical cost of giving up what's intuitive. We do not mean to suggest that these approaches are insufficiently developed, or that their proponents have in any way neglected their philosophical duties. On the contrary, each of the approaches we survey below are complex, interesting, and worth taking seriously. What they have in common, though, is their insistence that once we attend to some previously neglected factor, we will see that the threat from determinism dissipates. We call these compatibilisms "easy" as a simple way to demarcate them from approaches that address the worries about giving up the fixity of the past/laws head on.

But, as we will argue in this article, these apparently "easy" routes to compatibilism do not succeed. In particular, we will argue that these authors do not take all of the implications of determinism into account in developing their compatibilist positions. More specifically, all of these authors implicitly require us to give up either the fixity of the past or the fixity of the laws, but the authors do not make this explicit, nor do they explain how their views are preferable to traditional compatibilist views with respect to giving up the fixity of the past/laws. In addition to

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

highlighting a common thread throughout these apparently disparate approaches to free will, then, the result will be (we hope) new insight into what is misguided about this approach to arguing for compatibilism.

We will proceed as follows. In section 2, we will clarify the alleged threat to free will from determinism, focusing on Peter van Inwagen's presentation of the argument for incompatibilism. In sections 3–5, we will summarize and criticize the approaches of List, Sehon, and Markosian. We conclude, in section 6, with a discussion of an important lesson that may be learned from our survey of these novel and interesting, even if ultimately unsuccessful, approaches to compatibilism.

2. The Threat from Determinism

We began with a rough sketch of why someone might think that determinism and free will are incompatible, but it is worth spelling out the argument in detail and seeing what exactly it is about determinism that is thought to preclude free will. The most widely discussed presentation of the "Consequence Argument" for incompatibilism is van Inwagen's third version (1983, chapter 3). This presentation of the argument relies on a "no choice" operator *N*, where "N*p*" abbreviates "*p* and no one has, or ever had, any choice about whether *p*," as well as the following two inference rules (where "□" symbolizes logical necessity and "⊃" is the material conditional):

Rule Alpha: From □*p*, we may infer N*p*.

Rule Beta: From N*p* and N(*p*⊃*q*), we may infer N*q*.

Now let "P" abbreviate some proposition describing the entire intrinsic state of the world in the distant past, before there were any human beings, let "L" abbreviate the conjunction of the laws of nature, and let "Q" be a true proposition describing some human behavior. Here is the formal presentation of the argument:

1.	□((P&L)⊃Q)	definition of determinism
2.	□(P⊃(L⊃Q))	from 1, by modal and sentential logic
3.	N(P⊃(L⊃Q))	from 2, by Rule Alpha
4.	NP	premise, fixity of the past
5.	N(L⊃Q)	from 3 and 4, by Rule Beta
6.	NL	premise, fixity of the laws
7.	NQ	from 5 and 6, by Rule Beta

And since we may substitute any proposition describing human behavior for Q, the argument clearly generalizes. It follows that, if determinism is true, then we never have a choice about what we do, which is to say that we lack free will in the sense of the freedom to do otherwise.

This version of the Consequence Argument has been widely discussed, and many have been convinced of its soundness. 1 and 2 are implications of determinism

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

(at least as it is understood by philosophers working on these issues), and the two inference rules the argument employs are eminently plausible, though of course not universally accepted.[1] To accept this much and to nevertheless maintain compatibilism, then, would require giving up either the fixity of the past or the fixity of the laws (or both). Following Fischer's (1994) very helpful presentation of the dialectic, let us consider each option in turn.

Multiple Pasts Compatibilists (MPCs) give up the fixity of the past.[2] According to them, we need not hold fixed all facts about the actual past up to the time of an agent's action when assessing whether the agent is free to do otherwise than that action. For example, suppose it is causally determined that you read this sentence now. MPCs say that, nevertheless, you may well have been free to refrain from that action despite the fact that your doing otherwise would require the past (relative to the time of your action) to have been different. This is not to say that you have the ability *to change the past*—we are not imagining that the world in which you do otherwise shares the actual world's past up to the time of action and then (somehow) has a different past once you perform your action; rather, the world in which you do otherwise simply would have had a different past.

Local Miracle Compatibilists (LMCs) give up the fixity of the laws.[3] According to them, we need not hold fixed the laws of nature when assessing whether an agent is free to do otherwise than what the agent actually does. Again, suppose it is causally determined that you read this sentence now. LMCs say that, nevertheless, you may well have been free to refrain from that action despite the fact that your doing otherwise would require the laws of nature to have been different. This is not to say that you have the ability *to break the laws of nature*—we are not imagining that the world in which you do otherwise shares the actual world's laws of nature; rather, the world in which you do otherwise simply would have had different laws.

Neither of these versions of compatibilism will be the focus of this article, but it is worth noting that taking either of these approaches comes with a theoretical price tag. After all, many of us take our freedom at a given time to be constrained by what has happened earlier, as well as by what the actual laws of nature are. Even if it's true that the past or laws would have (had to have) been different had we acted otherwise, the fact of the past's being what is and the laws being what they are seems to limit what alternatives are genuinely available to us. To maintain that the freedom to do otherwise is compatible with determinism is to bite the bullet and endorse the counterintuitive. And it is exactly these counterintuitive implications of compatibilism that several recent approaches to compatibilism attempt to avoid.

[1] There has been considerable debate over the soundness of Rule Beta, but McKay and Johnson (1996) present an important counterexample, which has led van Inwagen (2000) to revise the argument. While Rule Alpha has received less attention, Spencer (2017) attempts to provide a counterexample to it as well.

[2] For example, see Saunders (1968), Perry (2004), and Dorr (2016).

[3] For example, see Lewis (1981), Graham (2008), Pendergraft (2011), and Tognazzini (2016). It is worth noting that Lewis's LMC requires giving up the fixity of the past as well as the fixity of the laws, which many will find counterintuitive though Lewis himself did not think that taking this approach required accepting anything counterintuitive.

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

3. Christian List's "Compatibilist Libertarianism"

Christian List (2019b) sets out to defend the reality of the traditional conception of free will—where that freedom requires that agents have, on at least some occasions, alternative possibilities—from certain scientific challenges. List makes his targets clear from the start, quoting passages from Sam Harris and Jerry Coyne, both of whom are convinced that science has disproved free will. Rather than tread the traditional compatibilist path, however, List attempts to play by the rules of scientific discourse, proffering a hypothesis which meets the "falsifiability requirement" demanded of all scientific theories.[4]

Key to List's view is the idea that different types of scientifically observed phenomena are described by different levels of disciplinary discourse, none of which is made obsolete by the other disciplines. In addition to "lower-level" domains like physics and chemistry, there are "higher-level" domains like economics and psychology, and we cannot dispense with the latter type of domain simply by doing more work in the lower-level domains, even if it turns out that the higher-level phenomena supervene on the lower-level phenomena. Accordingly, it may be that our best theories in physics turn out to be deterministic, and yet the laws governing the domains relevant to human agency (e.g., human psychology) may nevertheless be indeterministic. So free will, while it requires indeterminacy at one level (and thus List wants to call his view "libertarian"), it is compatible with determinism at the level of physics (and thus List wants to call his view "compatibilist"). Before we say more about List's "compatibilist libertarianism," or "free will emergentism," however, it is worth saying a bit more about List's view of the relationship between higher- and lower-level domains.

Even within the nuanced purview of the physical sciences, List suggests that there is a certain incommensurable hierarchy of supervenience. For example, the biologist is not going to turn to the physicist's description of kinetic energy to ascertain a pattern in amphibian mating practices. The type of description the biologist needs—the set of facts relevant to the object of inquiry—will not be decipherable from the gargantuan record of energy transfers at the most fundamentally observable level of matter in the frog. List claims that more abstract and broader descriptions will be necessary which are both unique to the discipline in question and not reducible to a more elementary, lower-level discipline. As List places the agential or psychological level atop his hierarchy, he admits that agential facts are irreducibly indispensable explanatory features of the human and social sciences. The principle which guides List in keeping his levels straight is what he deems the "naturalistic ontological attitude"—that is, a phenomenon can be said to exist if it is explanatorily indispensable to the best scientific theories within a given discipline. In this sense, free will can be said to exist given its necessity for making the human sciences work in the same way that quarks are said to exist as a necessary explanatory building block for theories in quantum mechanics.

[4] For an excellent critical summary of List (2019b), see Kaiserman and Kodsi (2021). The concern we raise for List's route to compatibilism is distinct from the worries raised there, as well as distinct from the issues raised by Mele (2020).

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

On List's view, free will requires alternative possibilities, but it may be that alternative possibilities are indispensable from the relevant higher-level domain of human psychology despite determinism at the lower-level domain of physics. According to List, the problem posed by determinism is that "given the initial state of the universe, only one course of events will have been physically possible" (2019b: 80). But, according to List, to infer from physical determinism that there is only one course of events that is *agentially* possible is to make a category mistake. As long as we accept that higher-level phenomena *supervene on* and *are multiply realizable in* lower-level phenomena, it will be possible for one level to be deterministic and the other indeterministic. List's claim about the relationship between levels, which we can call *supervenience with multiple realizability* (SMR), says not only that "[t]here cannot be variations in the agential state without variations in the physical state" but also that "not every variation in the physical state needs to bring about a variation in the agential state" (2014: 162; cf. 2019b: 91). And, List claims, so long as an agential state supervenes on a physical state but is multiply realizable in various physical states, we can have agential indeterminism even if the world is deterministic (at the level of physics).

But how exactly is this supposed to work? List provides some diagrams that help to show what he has in mind, and what follows is our simplified gloss, which will suffice for our purposes.[5] List creates a model of world histories at two different levels of description: the level of physics and the level of agency. Consider two times, t1 and t2. Assuming *supervenience with multiple realizability* (SMR), it could be that at one time, t1, an agential state A that actually obtains in the physical state P-t1 could have been realized by some non-actual physical state P*-t1. In addition, it could be the case that P-t1 is deterministically connected to some physical state P-t2, while P*-t1 is deterministically connected to P*-t2, and that two distinct agential states, B and B*, are realized by P-t2 and P*t2, respectively, at the later time t2, despite the very same agential state A having been realized in both P-t1 and P*-t1. In this model, List wants to say, there would be a deterministic connection between the physical states at t1 and t2 (between P-t1 and P-t2 and between P*-t1 and P*-t2), but an *in*deterministic connection between the *agential* states at t1 and t2 (A and B/B*).

The upshot of this model is that the world could be deterministic at the level of physics and yet *in*deterministic at the level of agency—i.e., the level relevant to free will—given SMR. List takes this to show that the Consequence Argument is unsound.[6] The Consequence Argument aims to show that determinism would imply that we never have alternative possibilities, given the fixity of the past and the fixity of the laws. List takes the Consequence Argument to be about the laws of physics and the state of the world in the past *at the level of physics*, so the only way for this threat from determinism to undermine free will would be if the following "linking thesis" were true: "*If*, given the complete physical state of the world at any point in time, only one future sequence of events is physically possible, *then*, in

[5] See List's Figures 1 and 2 on (2014: 166) and (2019b: 94).

[6] List (2014; 2019b) develops the approach in detail, and List (2019a) attempts to show how the approach undermines the Consequence Argument. Since the approach is spelled out in most detail in List (2019b), we will continue to refer primarily to that work.

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

any situation, only one course of action is ever possible for an agent" (2019b: 88). But, given SMR, "the totality of facts at the psychological level up to a given time may leave more than one future course of action open for an agent…so, there may be agential indeterminism, even in the presence of physical determinism" (2019b: 92), and thus List takes the linking thesis to be false (and the Consequence Argument unsound).

As we see it, List does not provide sufficiently good reason to deny the linking thesis, and this can be seen by thinking more carefully about the nature of the threat to free will from determinism codified in the Consequence Argument. What is problematic about List's move is that, given the supervenience of the agential level on the physical, if we hold fixed all of the *physical* facts about the past (and hold fixed the laws), only one future will be possible. In order to secure alternative possibilities at the level of agency, as he intends to do, what List needs is a restricted set of facts, one that doesn't hold fixed the past and laws (but only certain more "local" features of agents and their environments that are relevant to the agential sciences—psychology, decision theory, etc.). But List does not explicitly say that he's advocating for restricting the set of facts to be held fixed when assessing agents' freedom. Doing so would require rejecting either the fixity of the past or the fixity of the laws, and not only would this take us back to well-worn territory from the traditional compatibilist's response to the Consequence Argument, but List does not offer reasons for thinking we can reject either of these principles with impunity, and so it's not clear how his view is supposed to fare any better than traditional compatibilism.

Before turning to another way to put our criticism, it is worth pausing for a moment to address a claim List makes about the Consequence Argument. List (2019a) says that the Consequence Argument "involves a category mistake" since "it conflates two different levels of description, especially by placing physical-level propositions within the scope of agential-level modal operators" (2019a: 271–272). To be sure, the Consequence Argument is not typically formulated to take the physical- vs. agential-level distinction into account, and so it may seem as though the argument makes a category mistake. But, again, given the supervenience of the agential on the physical, this does not obviate the basic worry. We can run the argument using only physical-level propositions—e.g., replacing "Q," which is about some human behavior, with a physical-level proposition about any time at which a human being performs an action—but those physical-level propositions will imply agential-level propositions, and if no one has a choice about the truth of the relevant physical-level proposition, then no one has a choice about the supervenient agential-level proposition either.

Moreover, there is reason to think that there is nothing incoherent in making the claim that an agent has no choice about the truth of a physical-level proposition. For consider a non-agential proposition about medium-sized dry goods—that there's beer in the fridge, say (ignoring the fact that beer isn't dry). Surely the claim that no one has a choice about that proposition is coherent. But then so too should be a claim about an agent's having no choice about a physical-level proposition—that there's an electron in the CERN collider, say. It very well may be that an agent has a choice about that (perhaps their pressing a button caused the electron to be there). And if

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

there's no incoherence in having a physical-level proposition within the scope of an agential-level modal operator (like "N," which concerns what an agent has a choice about), then there's no reason to think that the Consequence Argument's use of such an operator involves a category mistake.

Here is another way to put our main criticism. It may well be that agential state A could have been realized in some other, non-actual physical state, and it may be that this would have deterministically led to an alternative physical state at a later time which would have realized a different, non-actual agential state. But A was *not* realized by that alternative physical state at the earlier time, and so there was no *genuine* alternative to the actual agential state at the later time to which A led. If we were to hold fixed a *restricted* set of facts (perhaps those relevant to the field of psychology), perhaps we could not predict which later agential state would be realized. But this is only an epistemic limitation on our part (and on the part of the agential level of description).

It is worth noting, before moving on, that List gives several hints that he thinks that the reality of free will may result from our epistemic limitations. In the introduction to his book, he says:

> Free will, I will argue, is in the company of other phenomena that emerge from the physical world but that are not best *understood* in fundamental physical terms themselves…All these ultimately emerge from physical processes, but we need to go beyond physics to *understand* them. Looking at them solely through the lens of the physical laws governing particles and molecules, for instance, *would give us little insight*. (2019b: 4, emphases added)

And then later, in his discussion of the ineliminability of alternative possibilities from the higher-level sciences of psychology and decision theory, List says: "It seems, then, that the assumption of agential indeterminism is at the heart of intentional explanation: there must be a sense of possibility in which an agent faces choices among several possible options; the goal is to identify the agent's chosen options among the possible ones" (2019b: 101). But, of course, the sciences in question require only epistemic possibilities, not genuine (metaphysical) possibilities. List is aware of this worry and addresses it briefly (2019b: 102–103), but he finds it more complicated and revisionary than "the more literal interpretation according to which the possible options are the possibilities the agent is genuinely faced with" (2019b: 103). We do not find List's brief remarks here persuasive, but we will set this aside for present purposes and return to our main challenge for List's view.

Our challenge for List may seem at first to be trivial, one which List has been attempting to address all along by claiming the insignificance of physical determinism upon agential freedom. Nevertheless, the problem for List's account of free will may be highlighted using an example from his own description of irreducible supervenience earlier in the book. Here, List provides the analogy of the relationship between a complete image and its constituent parts, whether particles or pixels (2019b: 73). On the physical level of description, the image is composed of an incredibly large and complex system of particles positioned in

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

certain ways and reflecting light in certain ways, but none of which, in their particularity, admit the complete image within themselves. A single atom in the smile of Mona Lisa, though it may be among the material constituents of the color of Mona Lisa's smile at the macrophysical level of description, does not within itself contain any of the "holistic" features of the painting. List claims that the interplay of the agential and physical levels are similarly related: just as the image of the Mona Lisa supervenes, but is not reducible to, the state of the atoms, agential states supervene on, but are not reducible to, physical brain states (or other relevant physical states). Despite some methodological concerns one might have with this analogy, we will grant it to List and use it to argue that List has not succeeded in showing that we should deny the linking thesis.

If determinism is true, then there is only one physically possible future—only one physically possible foliation of future events from the actual past, given the laws. In the case of the image, if we hold fixed the state of the particles within the image, there is only one possible rendering of the image, so to speak, given the supervenience of the image on the particles. It may be that different particles, or a different state of the same particles, would result in the same image, but holding fixed the actual state of the particles, a determinate image results. By analogy, for every physical state of a deterministic universe, there is only one possible rendering of the agential level. It may be that, for some description of human behavior at the agential level, the very same agential state could have been realized by different physical states, but holding fixed the actual physical state of the universe, a determinate agential state results. But if the physical state at one time is deterministically connected to the physical state at another time, and if each of these physical states results in a determinate agential state, then alternative agential states are not genuinely available. Even if we accept that free will is an emergent phenomenon, then, the fact of its supervenience on the physical implies that we cannot have indeterminacy at the agential level if determinism is true at the physical level.

One might object that we are simply violating List's injunction to keep our discussion of free will at the level of psychology or begging the question against List. After all, it may seem as though our argument simply insists that the physical constrains the psychological when that is precisely what List denies. But it is not question-begging for us to reject List's injunction. This is because List's claims about the supervenience of the psychological on the physical are in tension with his view that determinism (a thesis about the physical) is no threat to free will (a psychological matter). Given List's own claims about the supervenience of the psychological on the physical, we have reason to reject the injunction to limit our discussion of free will to the level of psychology. Our analogies are meant to elucidate how one level may bear on another when, even though distinct levels, one supervenes on the other.

One further analogy may be helpful to consider. Suppose you are playing a simple video game that, it turns out, is designed so that game states supervene on and are multiply realizable in bits of code. So, at an earlier time, t1, you are in game state A, which it turns out is realized by a particular string of code, C-t1, but might have also been realized by C*-t1. It seems to you, as you are playing the game, that as though one of two game states, B or B*, are both possible to bring about at a later time, t2. However, as it turns out, B would have to be realized by C-t2, and B* by C*-t2, and

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

code state C-t1 deterministically causes C-t2, whereas C*-t1 deterministically causes C*-t2. Were you to discover these details about the game while in game state A at t1, you would be right to think that you have no genuine alternative to B at t2. But, as we hope has been clear in our presentation of the case, this video game scenario is exactly analogous to List's model discussed above, and so we should think that List's model does not secure genuine alternative possibilities.

While we find much to admire in List's approach to compatibilism, especially its focus on the relationship between the various scientific domains, we think that it ultimately fails to address the threat from determinism. To do so, the approach would need to be supplemented with an account of why we should restrict the set of facts we hold fixed when assessing agents' freedom. Even so, such a restriction will require giving up either the fixity of the past or the fixity of the laws (or both). If both the past and laws are held fixed (and the world is deterministic), there can be only one foliation of future events, including human behavior. Thus, if we are to secure alternative possibilities at the level of human agency, then, when assessing agents' freedom, we must not hold fixed both the past and the laws. And so, even if List is trying to avoid this result, his approach requires giving up the fixity of the past or fixity of the laws.

Even if List's approach were supplemented with a denial of the fixity of the past or the fixity of the laws, though, such supplementation would be doing the heavy lifting in objecting to the Consequence Argument and so would take us back to well-worn traditional compatibilist territory. Not only would this approach fail to be an "easy" route to compatibilism, then, but it would also seem to fail to advance the debate. It may be, however, that there's a way to retool List's approach such that it does advance the debate. We will return to that suggestion in the conclusion, after considering two other attempts to establish compatibilism more easily.

4. Scott Sehon's Non-Causal Compatibilism

Scott Sehon's (2012; 2016) non-causal, compatibilist project has much in common with List's project as an attempt to explain free will through the lens of common-sense psychology as the irreducible property of a free agent. For Sehon's compatibilist theory, this irreducibility takes the form of a teleological, non-causal account of action explanation, which denies that action explanation can be reduced to causal explanations of human action: "on the teleological account of action explanation, the explanation does not work by citing an antecedent cause. Teleological explanations explain by citing a state of affairs or goal towards which the behavior is directed" (2012: 2). To use one of Sehon's examples, the explanation that "Jane is going to the cafe to meet her friend" can be construed in two ways. According to the causal account, on the one hand, some of Jane's mental states (e.g., her desire to meet her friend and her belief that she could meet her friend by going to the cafe) cause her behavior, and the explanation of her behavior works by citing the antecedent cause. According to the teleological account, on the other hand, Jane's behavior is explained by the goal toward which the behavior is directed, namely the goal of meeting her friend. Such teleological explanations, Sehon claims, are not reducible to causal explanations but rather are "irreducibly teleological" (2012: 3).

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

So far this might seem like a debate about action explanation that is orthogonal to the free will debate, but Sehon makes the striking claim that "free actions = actions for which we are responsible = intentional actions = goal-directed actions" (2016: 129), and each of these comes in degrees. In other words, Sehon thinks that "the behaviors that are teleologically explicable are the free actions, the ones for which we are responsible" (2016: 129), and so the more rationalizable an action, the more free it is, and the more responsible a person is for it, and vice versa. For us to have free will, then, is just for our actions to be teleologically explicable.

According to Sehon's view, we generate and justify a teleological explanation for an agent's action by constructing a theory of the agent, and the theory is constrained by the following principle (R): "Given two theories of an agent, it is unreasonable to believe one according to which the agent is significantly less rational" (2012: 4; 2016: 27). If there are two explanations for Jane's trip to the cafe, the reasonable teleological explanation for her behavior is the one according to which Jane is more rational. An initial but major challenge for Sehon's project is Davidson's challenge to the non-causalist (see Davidson 1963: 690–693), but we will assume, for the sake of argument, that Sehon's account can meet Davidson's challenge. Sehon goes on to say that "roughly put, we judge candidate explanations on two axes: the degree to which the explanation makes the behavior appropriate for achieving the goal, and the degree to which the goal is of value" (2016: 28). But, according to Sehon, we make such explanations and justify them without relying on causal explanations: "The net lesson is this: the basic way in which we in fact approach giving reason explanations of human behavior strongly indicates that the mode of explanation is not causal, but *sui generis*—irreducibly teleological" (2016: 89). Thus, whether an agent acts with free will is ultimately a matter of whether (and the degree to which) the agent's action is goal-directed.

But since he takes the threat to free will from determinism to be a worry about the causes of our behavior and not a worry for the teleological explanations of agents' actions, Sehon takes his position in action theory to allow for simple compatibilist responses to arguments for incompatibilism. We will return to the Consequence Argument in a moment, but Sehon's strategy, and our objection to it, is perhaps clearest with respect to an incompatibilist argument that Sehon calls the "Completely Fixed" Argument:

(1) My behaviors are completely fixed by the laws of nature and events in the distant past. [Premise-determinism]
(2) Laws of nature and events in the distant past are not under my control. [Premise]
(3) My behaviors are completely fixed by circumstances that are not under my control. [1,2]
(4) If a behavior is completely fixed by circumstances that are not under my control, then the behavior is not under my control. [Premise]
(5) If a behavior is not under my control, then it is not free. [Premise]
(6) If a behavior is completely fixed by circumstances that are not under my control, then it is not free. [4,5]
(7) My behaviors are not free. [3,6] (2016: 177; cf. Sehon 2012: 357)

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

This is a fair sketch of the general challenge from incompatibilism—one that highlights the intuitive threat to free will from determinism.

Sehon's response to this argument is to distinguish between two senses of the expression 'x is completely fixed by y' and then to argue that, on either disambiguation of the expression, either premise (1) or premise (4) will turn out to be false. Here are the two readings of the expression:

> Reading (A): x is completely fixed by y = y *completely* explains x.
>
> Reading (B): x is completely fixed by y = y *causally* explains x. (2016: 177–178)

If we go with (A), Sehon says, then premise (1) is false, since "there is also an explanation of my behavior, a teleological explanation, that doesn't refer to events in the distant past" (2016: 178). But if we go with (B), then Sehon takes premise (4) to be problematic, since he takes it that an action's being explicable on the basis of an agent's reasons suffices for the action to be under the agent's control, whereas the action's causal history is irrelevant to "its status as purposive or not" (2016: 178).

Sehon goes on to consider the Consequence Argument, and his response here is complicated by the fact that Sehon takes the argument to rely on a problematic characterization of determinism. Setting that issue aside, though, Sehon's response to a different "reading" of the argument employs the same strategy as his response to the Completely Fixed Argument. As before, let "N*p*" abbreviate "*p* and no one has, or ever had, any choice about whether *p*," but let's simplify by letting "'GS' represent a sentence expressing the global state of the universe, including its laws of nature, at some point in the distant past before any human beings existed" (2016: 179) and by letting "'P' represent any true sentence, for example, 'Scott had coffee with breakfast on 1 February 2015'" (2016: 179). Now consider the following argument and inference rules:

> (1) (GS explains P) [Premise—determinism]
> (2) N(GS explains P) [Application of rule α to (1)]
> (3) N(GS) [Premise]
> (4) N(P) [application of rule β to 2 and 3)
>
> The required rules of inference:
> rule (α) from *p* explains *q* infer N(*p* explains *q*)
> rule (β) from N(*p explains q*) and N*p*, infer N*q* (2016: 182)

Sehon's response to this argument is exactly the same as his response to the Completely Fixed Argument: he argues that it trades on an equivocation on the expression 'x is completely fixed by y'. In this case, if we go with reading (A), premise (1) is false, since the global state of the universe does not answer explanatory questions about the purpose of the behavior specified in P. But if we go with reading (B), rule (β) may be called into question, for "whether a behavior counts as an action is not determined by the nature of its causal history" (2016: 183), and so "the causal conditions specified in the antecedent have no obvious relevance to the

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

consequent's conclusion that [the behavior specified in P is not] a free action" (2016: 183).

To see where we think Sehon's compatibilist strategy goes wrong, it will be helpful to consider Sehon's discussion of Alfred Mele's "recherché hypothetical" (Sehon 2016: 57; cf. Mele 2003; 2010) about Norm and the Martians. In Mele's case, an ordinary human agent Norm is taken over by Martians who will shoot "M-rays" at him when he is about to actualize an intention or desire to do something, and the M-rays completely control all of his physical functions in such a way that Norm cannot even tell that he is being "manipulated" by something beyond his own intent. However, the Martians are to relinquish control as soon as Norm alters his intentions or desires. The effect is that Norm will still do exactly as he pleases, but the physical movements required to actualize an intention will be handled by the Martians' M-rays. In the cases where Norm is being controlled by the M-rays, it appears that Norm is not acting, yet his behavior is rationalizable in just the way that Sehon thinks suffices for Norm's behavior to count as an action (and a free one, at that). (Sehon 2016: 57) Cases like this reveal an important feature of Sehon's position: the teleological realist holds only extremely local facts fixed when considering agency. Sehon admits as much when he writes: "Since teleological explanations explain by citing a state of affairs or goal toward which the behavior is directed, such explanations need not cite any antecedent mental state of the agent at all" (2016: 25). Norm can be said to have agency—including free will—in Sehon's account because his behavior is rationalizable, and this has nothing to do with the way in which the behavior was (causally) brought about.

But even granting that Norm is *acting* when he is manipulated in this way by the Martians, it seems clear that, given the M-rays, Norm is bound to act in exactly the way that he does, given his intentions and desires—i.e., he lacks the freedom to do otherwise at the time of action. It may be that, had Norm intended otherwise or desired differently, then he would have done otherwise (as a result of the Martians using their M-rays differently). And it may be that, in that case, there would be an alternative teleological explanation of Norm's behavior than the one that fits in the actual case. What seems clear, though, is that Norm lacks the freedom to do otherwise, and the fact that there *could* have been some alternative teleological explanation of his behavior is neither here nor there.

And now return to the threat to free will from determinism. If determinism is true, and if we hold fixed the past and the laws when assessing agents' abilities, only one teleological explanation will be possible, and so there won't be alternative possibilities. True, as Sehon says in reply to the Consequence Argument, the global state of the world in the distant past may not explain why an agent acted as they did today; still, an implication of determinism is that the fact that the agent acted as they did today was entailed by facts about the global state of the world in the distant past. In order to secure a genuine freedom to do otherwise, then, we would need to hold fixed only a restricted set of facts—one that doesn't hold fixed the past and laws (but only certain more "local" features of the agent and their environment that are relevant to teleological explanation). But Sehon does not seem to take himself to be giving up either the fixity of the past or the fixity of the laws (this is not explicit, if he does), nor does he offer reasons for thinking we can reject either of these

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

principles with impunity. Thus, it's not clear how this is any better than traditional compatibilism.[7]

5. Ned Markosian's Agent-Causal Compatibilism

Ned Markosian (1999; 2012) provides the third and final novel approach to compatibilism that we will survey. Like the previous two compatibilist theories of List and Sehon, Markosian (2012) attempts to locate agency as an irreducible phenomenon removed from event-causal chains by proffering agent-causation as a solution. The specific form that Markosian's agent-causation assumes is expressed in The Compatibilist Version of the Theory of Agent Causation (COMTAC), which claims, "A is *morally free* iff A is caused by A's agent" (2012: 384). Though Markosian notes that agent-causalists are frequently incompatibilists, his specific pitch in Markosian (2012) is to recruit event-causal compatibilists to the agent-causalist camp. The distinction between these causal theories is simple enough: event-causalists maintain that an agent's causal contribution to an action is reducible to causation by events, whereas agent-causalists maintain that causation by agents—typically conceived as causation by agents-as-substances—is not reducible to causation by events. Markosian uses this specific expression of agent causation to leverage a compatibilist response to the most salient challenges from incompatibilism.

The main thrust of Markosian's formulation of agent causation claims that the requirements of moral freedom are met even in the absence of indeterminism in the causal history of an action, if it is caused by an agent. To continue the example which Markosian uses to illustrate this thesis, suppose that there is someone named Yasmin who asks Imran to pass her the salt, and let "e1" stand for the action of Yasmin's asking. Imran obliges and passes the salt (call this "e2"), to which Yasmin responds with thanks ("e3"). How would such a sequence of events be causally rendered? For Markosian, there may be a sequence of causally determined events from e1 to e3, which nevertheless admit a distinct, second cause of e2 which is not merely reducible to the events that caused Imran's passing the salt. This is the agential causation which Imran affects, located solely in the mental states and intentions of Imran in his act of passing the salt. Markosian does not delve into any deeper account of the function of this species of causation, but it suffices to say that the relevant feature of the picture is that agent causation irreducibly belongs to the agent who affects a morally free action. Agent causation does not preclude coexistent event causation, but exists as an independent causal source of an event. In Markosian's rendering of the Imran

[7] One might object that we are simply violating Sehon's anti-reductionism in citing lower-level facts (about the past and laws) as a reason to deny that multiple teleological action explanations are available, if determinism is true. (Thanks to an anonymous reviewer for raising this concern.) But, as we said about our response to List, here too we see ourselves as giving a reason for thinking that we should not keep our discussion only at the higher level (in this case, the level of teleological explanations). If determinism is true, then lower-level facts about the past and laws entail facts about the world today, including about everything that would be relevant for teleological explanations of agents' actions. If we ignore the lower level facts, then, we may mistakenly think that different teleological action explanations were available.

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

anecdote, Imran may be said to be both the morally free, agential cause of e2, just as much as e2 is a member of an event-causal chain preceded by e1.

Like the previous evaluations of List's and Sehon's compatibilist theories, the strength of Markosian's agent-causation compatibilism must be evaluated on the basis of its addition to the compatibilist-incompatibilist debate. Though he deals rather summarily on the issue, it is in Markosian's treatment of the Consequence Argument that the weakness of agent-causal compatibilism is most evident. Simply put, Markosian wants to say that the threat to free will from determinism, codified in the Consequence Argument, disappears when we consider agent causation. Considering the brevity of Markosian's response in (2012), we quote the passage in full:

> Now let me sketch a variation on this response to the modal argument that is available to the COMTACer. The COMTACer can say that (β) is invalid precisely because of cases involving agent causation. Here is the idea. Suppose it was not up to you either what was going on a million years ago, or whether what was going on a million years ago determined what you will do next. Still, if we assume that you have the power of agent causation, and that you will literally be a cause of your next action, then it's very plausible (I think) to say that it *is* partly up to you what you will do next. After all, on this assumption, there is a causal sequence that is initiated by you and that leads to your next action. Hence it is partly up to you what you will do next in the sense that you personally are among the causes of what you will do next. (2012: 393)

The crux of the worry for Markosian is that, if we assume that an agent's action is part of a deterministic event-causal sequence, and if we assume the fixity of the past and laws, then only that very action will be possible for the agent whether it is agent-caused or not. In other words, an implication of determinism is that facts about what we will do are entailed by facts about the past and the laws of nature, so if we hold fixed these facts about the past and laws, then only one course of action will be possible. Thus, determinism poses the same threat to our freedom to do otherwise whether what we will do is brought about simply by events (as the event-causalist maintains) or also (or even only) by agents-as-substances.

Why does Markosian think that "if we assume that you have the power of agent causation, and that you will literally be a cause of your next action, then it's very plausible (I think) to say that it *is* partly up to you what you will do next" (2012: 393)? If we are construing something's being "up to you" as requiring the freedom to do otherwise, it will only seem plausible that a determined agent's agent-causing an action makes it "up to them" if we are bracketing the facts about the events and laws of nature that entail what the agent will do. Again, as with List and Sehon, Markosian's brand of compatibilism only works if local causal facts are held fixed, while distant causal facts are not. Markosian's sense that adding agent-causation into the picture can secure free will presupposes that there are things which causally escape the purview of determinism. The incompatibilist's point is that, since the past and laws are fixed, anything entailed by the past and laws is not up to us. Given this, determinism implies that there can be no possible deviation from the

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

single future which those facts admit, even if the future includes causation by agents-as-substances. Markosian's account, then, does not reconcile the freedom to do otherwise with determinism, as traditional compatibilists aim to do.

Interestingly, Markosian (2012: 394) explicitly denies that moral freedom (and moral responsibility) requires the freedom (or ability) to do otherwise, indicating that he takes his proposal to be a source compatibilist proposal. We find it puzzling that this comes after Markosian's critical response to the Consequence Argument—a response that is unnecessary if the freedom to do otherwise is not required for moral responsibility, as that is the sense of "up to you" at stake in the Consequence Argument. Setting that aside, if we take Markosian to be offering a source compatibilist proposal, then it may seem that he is not offering an "easy" compatibilism after all; instead, he seems to be following well-worn source compatibilist ground.

But Markosian thinks that the compatibilist should adopt an agent-causal theory anyway, since "it's much easier to convince people that the ability to do otherwise is not a necessary condition for moral responsibility if you appeal to agent causation" (2012: 394). In support of this claim, Markosian returns to the case of Imran's passing the salt. Supposing that Imran cannot do otherwise than pass the salt, it may seem that he is not morally responsible for passing it. But if it is made explicit "that Imran exercises the power of agent causation, and causes himself to pass the salt" (2012: 394), Markosian says, then it is obvious (or at least very plausible) that he is morally responsible for passing it.

We do not think that the mere addition of agent causation to this story shows that Imran is morally responsible for passing the salt. Markosian claims that, whereas it would be dubious whether Imran is morally responsible for passing the salt if agent causation were not involved, it is obvious (or at least very plausible) that Imran *is* morally responsible for passing the salt when agent causation *is* involved. But Markosian does not provide any support for this claim. Without any support for this claim, it is hard to see why source compatibilists should accept it, let alone why someone convinced that moral responsibility requires the ability to do otherwise would be convinced. What this means is that, even if the compatibilist denies that moral responsibility requires the ability to do otherwise, their source compatibilism will have to be motivated on different grounds.[8] As such, it is difficult to see what advantages are offered to compatibilism by Markosian's agent-causalism.[9]

6. Conclusion: Where Do We Go from Here?

We have argued that several recent "easy" attempts to establish the compatibility of free will and determinism are not ultimately successful. Whether it is List's

[8] For example, they may appeal to "Frankfurt examples"—so named due to Frankfurt's (1969) influential discussion of them—to support the view that moral responsibility does not require the freedom to do otherwise, as source compatibilists like Fischer (2006) and Sartorio (2016) do. This is well-worn ground, and Markosian wants to avoid relying on these examples (2012: 394), but, again, it's not clear how the addition of agent causation is supposed to show that moral responsibility does not require the freedom to do otherwise.

[9] This is not to say that there are no reasons to endorse agent-causalism over event-causalism, but we agree with Franklin (2016) that any reason for such an endorsement will be orthogonal to the compatibilist-incompatibilist debate.

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press

compatibilist libertarianism, Sehon's non-causal compatibilism, or Markosian's compatibilist theory of agent causation, none of these approaches sufficiently appreciates the threat from determinism and the difficulty of preserving the freedom to do otherwise in a deterministic world. Each author presents an account of free will that assesses agents' freedom without holding fixed all of the facts about the past and laws of nature, but none of these authors recognizes their commitment to giving up either the fixity of the past or the fixity of the laws. Moreover, it is unclear to us how the considerations that these authors appeal to—the level of agency (as opposed to the level of physics), teleological action explanation, and agent causation—are going to provide an argument for giving up either the fixity of the past or the fixity of the laws.

That said, we do see an important lesson to be learned from our survey of these three approaches to compatibilism. *If* these approaches were to be supplemented with a new and principled reason for rejecting the familiar and time-honored principles of the fixity of the past and laws, this would allow for dialectical progress. For example, if List were to argue that, given some feature of the role of alternative possibilities at the higher-level domain of human psychology, it turns out that we have new reason to suspect that human behavior is not constrained by the actual past, this new reason for denying the fixity of the past would push the dialectic forward. Similarly, perhaps Sehon or Markosian might recognize their commitment to the denial of at least one of these principles and appeal to their work in traditional philosophy of action and the metaphysics of causation to call either the fixity of the past or the fixity of the laws into doubt. It is difficult for us to imagine exactly what this would look like in *substance*, but we think that, at least *structurally* speaking, this is the form that progress must take.

As it stands, however, traditional compatibilist positions (multiple-pasts compatibilism and local miracle compatibilism) avoid the problems we've raised for these new approaches, insofar as they recognize their denial of one of the intuitive principles (the principles of the fixity of the past and laws). But these positions aren't easy compatibilisms—they come with the dialectical cost of giving up what's intuitive. While we are not sanguine about the success of the project, one way to push the dialectic forward—mitigating the cost of traditional compatibilism—would be to develop arguments from the considerations we have surveyed to the conclusion that either the fixity of the past or the fixity of the laws should be abandoned.

TAYLOR W. CYR
PHILOSOPHY, SAMFORD UNIVERSITY, UNITED STATES
taylor.w.cyr@gmail.com

PARKER GILLEY
UNIVERSITY OF VIRGINIA LAW SCHOOL, UNITED STATES
pgilley@samford.edu

Acknowledgements. For discussion and comments on earlier versions of this article, we are grateful to audiences at an informal presentation at Samford University; the members of the "Volitions, Intentions and Mental Causation: Theoretical and Experimental Approach" workshop in November of 2020; and the members of the "Compatibilist Libertarianism: Advantages and

Challenges" workshop in October of 2021. Thanks also to several anonymous reviewers and the editor of this journal.

Statements and Declarations. This work was partially supported by an Alabama Power Grant for Samford University Fellows.

Competing Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Davidson, Donald. 1963. "Actions, Reasons, and Causes," *Journal of Philosophy* 60: 685–700.
Dorr, Cian. 2016. "Against Counterfactual Miracles," *Philosophical Review* 125: 241–286.
Fischer, John Martin. 1994. *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell.
Fischer, John Martin. 2006. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.
Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66: 829–839.
Franklin, Christopher Evan. 2016. "If Anyone Should Be an Agent-Causalist, then Everyone Should Be an Agent-Causalist," *Mind* 125: 1101–1131.
Ginet, Carl. 1990. *On Action*. Cambridge: Cambridge University Press.
Graham, Peter. 2008. "A Defense of Local Miracle Compatibilism," *Philosophical Studies* 140: 65–82.
Kaiserman, Alex and Kodsi, Daniel. 2021. "*Why Free Will Is Real*, by Christian List," *Mind* 130: 987–996.
Lewis, David. 1981. "Are We Free to Break the Laws?" *Theoria* 47: 113–121.
List, Christian. 2014. "Free Will, Determinism, and the Possibility of Doing Otherwise," *Noûs* 48: 156–178.
List, Christian. 2019a. "What's Wrong with the Consequence Argument: A Compatibilist Libertarian Response," *Proceedings of the Aristotelian Society* 119: 253–274.
List, Christian. 2019b. *Why Free Will Is Real*. Cambridge: Harvard University Press.
Markosian, Ned. 1999. "A Compatibilist Version of the Theory of Agent Causation," *Pacific Philosophical Quarterly* 80: 257–277.
Markosian, Ned. 2012. "Agent Causation as the Solution to All the Compatibilist's Problems," *Philosophical Studies* 157: 383–398.
McKay, Thomas and Johnson, David. 1996. "A Reconsideration of an Argument against Compatibilism," *Philosophical Topics* 24: 113–122.
Mele, Alfred. 2003. *Motivation and Agency*. New York: Oxford University Press.
Mele, Alfred. 2010. "Teleological Explanations of Actions: Anticausalism versus Causalism," in *Causing Human Actions: New Perspectives on the Causal Theory of Action*, eds. J. Aguilar and A. Buckareff (Cambridge, MA: MIT Press): 183–198.
Mele, Alfred. 2020. "Free Will and Luck: Compatibilism versus Incompatibilism," *The Monist* 103: 262–277.
Pendergraft, Garrett. 2011. "The Explanatory Power of Local Miracle Compatibilism," *Philosophical Studies* 156: 249–266.
Perry, John. 2004. "Compatibilist Options," in *Freedom and Determinism*, eds. J. Campbell, M. O'Rourke, and D. Shier (Cambridge, MA: MIT Press): 231–254.
Sartorio, Carolina. 2016. *Causation and Free Will*. New York: Oxford University Press.
Saunders, John Turk. 1968. "The Temptation of Powerlessness," *American Philosophical Quarterly* 5: 100–108.
Sehon, Scott. 2012. "Action Explanation and the Free Will Debate: How Incompatibilist Arguments Go Wrong," *Philosophical Issues* 22: 351–368.
Sehon, Scott. 2016. *Free Will and Action Explanation: A Non-Causal, Compatibilist Account*. Oxford: Oxford University Press.

Spencer, Jack. 2017. "Able to Do the Impossible," *Mind* 126: 466–497.
Tognazzini, Neal. 2016. "Free Will and Miracles," *Thought* 5: 236–238.
van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon.
van Inwagen, Peter. 2000. "Free Will Remains a Mystery," *Philosophical Perspectives* 14: 1–19.

https://doi.org/10.1017/apa.2025.10016 Published online by Cambridge University Press