

The Naturalistic Case for Free Will

Christian List

Pre-print of a contribution to Stavros Ioannidis et al. (eds.), *Levels of Reality in Science and Philosophy*, Springer

Abstract: The aim of this expository paper is to give an informal overview of a plausible naturalistic case for free will. I will describe what I take to be the main naturalistically motivated challenges for free will and respond to them by presenting an indispensability argument for free will. The argument supports the reality of free will as an emergent higher-level phenomenon. I will also explain why the resulting picture of free will does not conflict with the possibility that the fundamental laws of nature are deterministic, and I will address some common objections.

1. Introduction

Skepticism about free will has become ever more prominent in public discourse. If one browses through the popular science literature or follows social-media coverage of the topic, one is likely to come across plenty of writings suggesting that free will is an illusion: a left-over from an outmoded, pre-scientific way of thinking that has no place in modern science. Sam Harris (2012), Jerry Coyne (2014), and Yuval Noah Harari (2016) are just three well-known writers who have made such claims. The skeptics typically appeal to a materialist worldview in which there is no place for genuine human agency and cite neuroscientific studies allegedly showing that human actions are caused not by our intentional mental states, but by physical processes in the brain and body as well as external influences. More broadly, they ask, if everything in the universe is governed by the laws of nature, and our actions are part of that universe, then how could those actions be free? And how could we legitimately be held responsible for them?

Many free-will skeptics have a noble moral motive, alongside their conviction that science is on their side: they find the present criminal justice systems in many countries unjust and wish to argue for criminal justice reform. But one can agree on the need for an overhaul of our criminal justice systems while still thinking that we shouldn't throw the notion of free will out of the window. Accepting the reality of free will is compatible with advocating criminal justice reform and supporting a more rehabilitative and less retributivist approach to punishment. And independently of its relevance to criminal justice, the idea of free will is central to our human self-understanding as agents. How, for instance, could we genuinely deliberate about which course of action to take – say, when we choose a job, a partner, or a political cause we wish to endorse – if we didn't take ourselves to be free in making this choice?

In this paper, I will outline my strategy for defending free will against the growing scepticism. Crucially, I do not proceed by denying science or watering down the definition of free will. Rather, I suggest that if we understand the lessons of a scientific worldview correctly, the idea of free will – in a fairly robust sense – is not just consistent with such a worldview but supported by it. In short, there is a naturalistic case for free will. My defence of free will is developed more fully and precisely in my recent book, *Why Free Will is Real* (2019a). The paper can be read as an informal summary of, and guide to, this defence.

I will first describe what I take to be the main challenges for free will from a scientifically informed perspective and then outline my response. And I will illustrate my strategy by zooming in on the most widely discussed challenge: the challenge from determinism. Finally, I will address some common objections.

2. The challenge

Let me begin with the overall challenge. Free will can be defined, on a first approximation, as an agent's capacity to choose and control his or her own actions. Free-will skeptics argue that there is no room for this capacity in a universe in which everything is the result of physical processes. The challenge can be made more precise in terms of a general argument scheme. The skeptics typically assume that free will requires some precondition – call it property P – which might be one or perhaps all of the following:

- intentional, goal-directed agency,
- alternative possibilities among which we can choose, and
- causation of our actions by our mental states, especially by our intentions.

Then they claim that science shows that there is no such thing as property P. In particular, they argue that intentional agency, alternative possibilities, or mental causation cannot be found among the fundamental physical features of the world. Regardless of whether you consult particle physics, biochemistry, or even neuroscience, you won't get around the fact that human organisms are collections of physical building blocks, all of which are ultimately governed by the laws of physics. And this, it seems, leaves very little room for intentional agency, alternative possibilities, and causal control over our actions. For this reason, the skeptics say, property P – whichever one of the three it is – is at best a convenient fiction of our pre-scientific way of thinking. It is not an ingredient of our physical universe. And so, since property P is required for free will, there is no free will.

Different naturalistic arguments against free will target different substitution instances for P. Some claim that intentional agency is an illusion. Intentionality does not fit into the physical universe. The idea that humans are agents with goals and purposes is a remnant from folk psychology, to be replaced by a more mechanistic understanding of the human organism as a bio-physical machine. On this picture, the traditional psychological understanding of humans as intentional agents will ultimately be replaced by a more reductionistic, neuroscientific understanding. I call this the “challenge from radical materialism”.

A second set of arguments claim that if the fundamental laws of physics are deterministic, as in a mechanical clockwork, then human beings could never have any alternative possibilities to choose from. Any past state of the universe – say at the time of the Big Bang – would have been sufficient to determine everything that was going to happen thereafter. When I chose to have tea rather than coffee this morning, to give a trivial example, I could not have acted otherwise. My choice was fixed by the world's prior conditions, under the laws of nature, as was your choice to read this

paper. I call this the “challenge from determinism”. It is, by far, the most widely discussed challenge for free will.

A third set of arguments, finally, assert that it is illusory to think that our actions are caused by our intentions. When I act, it is my brain that makes me do it. Any consciously experienced mental state to which I might intuitively attribute my action is only an epiphenomenon accompanying the real, physical cause – a byproduct. I call this the “challenge from epiphenomenalism”.

Unless we are prepared to say that intentional agency, alternative possibilities, and mental causation are not all needed for free will, the success of even just one of these arguments poses a significant challenge for free will. And on the face of it, neither intentional agency, nor alternative possibilities, nor mental causation are easy to give up as conditions for free will. Entities that don’t qualify as intentional agents don’t even seem to be *candidates* for the ascription of free will. Similarly, entities that never face any real choices between different alternatives don’t seem to qualify as “free” either. To claim that there is free will without choices would require at least a revisionary reinterpretation of the notion of freedom. Finally, if an entity’s behaviour is caused not by any intentional, mental states, but by completely sub-intentional, physical processes, then whatever this entity does can hardly be attributed to its own free will.

Furthermore, although the popular-science versions of these skeptical arguments have received much attention in public discourse, there are more academic versions too. These include, but are not restricted to, Patricia and Paul Churchland’s neuroscientific arguments for “eliminativism” about intentional agency (1981, 1986), Peter van Inwagen’s “consequence argument” for the incompatibility of free will and determinism (1975), Jaegwon Kim’s “causal exclusion argument” against certain non-reductive forms of mental causation (1998), and Benjamin Libet’s and other scholars’ experimental results on the neuronal activity underlying voluntary motor actions (1983). In sum, there appears to be a strong naturalistic case against free will.

How should we respond? One response is to conclude that there is no free will. That’s what the skeptics say. I find that response unsatisfactory. My view is that we should abandon such a central tenet of our common-sense understanding of the human condition only if the arguments against it are truly compelling, and I don’t think they are, as I will explain.

A second response, given by many free-will compatibilists, is to insist that free will doesn’t require all of the things I have mentioned, or that it requires them only in a weaker form. For instance, one might say, it is not necessary for free will that we have alternative possibilities to choose from. What matters for free will is merely that we endorse the choices we make, not that we could have acted otherwise. Alternatively, one might redefine what it means to say that an agent “could have acted otherwise”. Instead of interpreting it to mean that it was *possible* for the agent to act otherwise, one might interpret it to mean that *if* the world had been a little different than it actually was – say, the agent had tried to do something other than what he or she did – then he or she would have succeeded. We might then be able to bypass some of the challenges I have summarized. I am not convinced by such a response either, because it arguably comes at the cost of watering down

the notion of free will. It's not clear that such a weakened notion can do all the work we expect the notion of free will to do, as a basis for our self-understanding as responsible agents capable of deliberating about what to do. I consider the idea that we sometimes face genuine forks in the road central to our sense of responsible agency.

My own response to the challenge is different. I concede the skeptics' starting point and accept that free will does indeed require intentional agency, alternative possibilities to choose from, and causal control over our actions. And I also concede that if we look at the world solely through the lens of fundamental physics or even that of neuroscience, we are unlikely to find agency, choice, and mental causation. But I argue that this does not show that these properties are unreal. Rather, free will and its prerequisites are emergent, higher-level phenomena. They emerge from physical processes but are not reducible to them. I will now explain this response in more detail.

3. Free will as a higher-level phenomenon

As noted, I accept that free will requires intentional agency, alternative possibilities among which we can choose, and causal control over our actions. I take these three properties, suitably understood, to be individually necessary for free will and jointly sufficient. That is, there is no free will without intentional agency, alternative possibilities, and mental causation, and conversely, if all three properties are present in an entity, this is enough for free will; nothing else is needed. Characterizing free will in this way has two advantages. First, it arguably captures a relatively robust common-sense understanding of free will, in line with the "libertarian" intuitions that many people have before they encounter free-will skepticism. Second, by disaggregating free will into three properties, it clarifies what is at stake in the debate. The difficult and often emotionally charged question of whether humans have free will is replaced by a set of more tractable and somewhat less charged questions: whether humans are intentional agents, whether they have alternative possibilities, and whether their actions are caused by their mental states. This gives us a checklist of things we need to consider if we wish to find out whether there is free will. (For an earlier, similar definition of free will in terms of three properties, see Walter 2001, p. 6.)

How, then, can we establish that humans have all three properties? The key point to note is that there are two very different ways in which we can think about human beings. We can either think of them as physical systems, consisting of gazillions of interacting particles, and insist that human behaviour is to be understood as nothing but a physical process. Or we can think of humans as not just physical but also psychological, as beings with mental states and cognitive processes that underpin their behaviour. Call the first way of thinking the "reductionistic" one, and the second the "non-reductionistic" one.

It should be clear that if we adopt the reductionistic way of thinking, we may not find support for intentional agency, alternative possibilities, and mental causation. Intentionality may not seem to be a feature of physical systems; alternative possibilities may seem to conflict with physical determinism; and mental causation seems to go against the principle that all physical events must be attributable to physical causes. So, the reductionistic way of thinking leads directly to the free-

will skepticism I have described. However, the human and social sciences – from anthropology and psychology to sociology and economics – support the non-reductionistic way of thinking, which represents humans not as mere physical systems, but as agents with goals and purposes, beliefs and desires, and explains human behaviour on that basis. It would be impossible to make sense of human behaviour in its breadth and richness if we did not understand humans in this way. And this understanding, in turn, vindicates agency, choice, and mental causation as central features of human beings – features that emerge from (and “supervene” on) physical processes in the brain and body but do not lend themselves to a reductionistic description in physical terms alone.

Let me give you an analogy. Suppose someone claims that there is no such thing as unemployment. Why? Because unemployment does not feature among the properties to which our best theories of fundamental physics refer. If you consult quantum mechanics, for instance, then you won’t see any unemployment. But it would be absurd to conclude from this that unemployment is unreal. It is very much a real phenomenon, albeit a social and economic as opposed to purely physical one. And of course, this verdict is supported by our best scientific theories at the relevant level, such as sociology and economics. Those theories recognize the reality of unemployment, and it features as an *explanans* and an *explanandum* in social-scientific explanations. Like the skeptic who mistakenly searches for unemployment at the level of quantum mechanics, the free-will skeptics, I argue, make the mistake of looking for free will at the wrong level, namely the physical or neurobiological one – a level at which it cannot be found.

Free will and its prerequisites – intentional agency, alternative possibilities, and mental causation – are in the company of other emergent phenomena, from organisms and ecosystems to economies. These phenomena, too, would be hard to see if we were to look at the world solely through the lens of (say) physics or chemistry. We would see only particles and molecules, fields and forces, but no organisms, ecosystems, and economies. They are irreducibly higher-level phenomena, but that makes them no less real.

4. Why are intentional agency, alternative possibilities, and mental causation explanatorily indispensable?

Let’s begin with intentional agency. However much the different human and social sciences – such as anthropology on one side and economics on the other – disagree about how to explain human behaviour, the one thing they all have in common is that they take what Daniel Dennett (1987) calls an “intentional stance” towards human beings. That is, they explain human beings as agents who perceive the world and cognitively represent it, who act in pursuit of goals, and who respond to their situation in ways that are at least partly rational. Whether you consult anthropology or micro-economics, psychology or sociology, you will find this intentional mode of explanation as a common feature. By contrast, if we tried to make sense of human beings solely as heaps of interacting particles, or as complicated neural networks, we would at most be able to explain some details of the brain and body or some specific aspects of physiology and cognition – for instance, how the visual cortex implements certain perceptual tasks. We would not be able to explain the rich patterns of human behaviour in their breadth and flexibility.

To give a simple example, if I ask a taxi driver to take me to Victoria Station on one day, and I ask another taxi driver to take me to King's Cross Station on the next day, and each time I successfully reach my destination, it would be extremely hard – perhaps impossible, in practice – to explain in purely physical terms what the two events have in common. We would have to cite the incredibly complicated neural and other physical processes in each driver's brain and body as well as in the car. Contrast this with the intentional mode of explanation. Once we recognize the two taxi drivers as intentional agents who understand where I wish to go, form the intention to drive me there, and have an intelligible reason to do so, we can easily explain what's going on and make predictions on that basis. The assumption that the drivers are intentional agents is vindicated by its explanatory success. Generally, the ascription of agency to people is indispensable for a satisfactory explanation of their behaviour. This point should be fairly uncontroversial.

Next consider alternative possibilities. Just as we wouldn't be able to explain human behaviour without recognizing people as agents, so we wouldn't get very far in explaining behaviour if we didn't view people as making choices in which alternative actions are open to them. The idea that humans face choices between different options, consider them (where this can take a variety of forms ranging from quick processing to slow and careful deliberation), and select one option among the possible ones is no less important for the human and social sciences than the idea of agency itself. This means that we represent humans not as deterministic machines, but as beings for whom different courses of action are possible. I call this idea "agential indeterminism". Even in a field like decision-and-game theory in economics, which is sometimes (mis)interpreted as representing humans as nothing more than utility-maximizing automata, the notion of a decision tree with choices between several possible options is central. Having different options does not mean that they are all equally likely to be chosen. After deliberation, a decision-maker may well find some options more rational or more attractive than others.

I argue in my book that the assumption of agential indeterminism is a key presupposition of the intentional mode of explanation itself. Without that assumption, our explanations of people's behaviour in the human and social sciences would not get off the ground. My conclusion here is similar to that reached by Helen Steward (2012), who argues that the very idea of agency requires some form of indeterminism. Now one may legitimately ask whether the required agential indeterminism doesn't conflict with physical determinism. As I will explain in the next section, agential indeterminism is compatible with physical determinism – an initially surprising point which, despite sounding counterintuitive, can be established in a formally precise way.

Finally, let's turn to mental causation. Skeptics argue that it is not our conscious intentions that cause our actions, but physical states of the brain. On the theoretical side, they cite Jaegwon Kim's "causal exclusion argument" (1998), which asserts that if we attribute our actions to anything other than a physical cause, this will breach some central tenets of a scientific worldview, such as the principle that there are no physical effects without physical causes or the principle that we should not postulate more causes than strictly necessary. If a physical cause, such as a neural state of my brain, suffices to account for the movement of my arm, for instance, then we should not postulate any further mental cause. On the empirical side, skeptics cite a series of experiments conducted by

Benjamin Libet and his co-authors (1983), and subsequently others, showing that, when subjects are asked to perform voluntary movements of their limbs, one can detect some preparatory brain activity – a neuronal readiness potential – before the subjects experience the conscious intention to act. Libet took this to show that our intentions are only passive byproducts of the real physical causes.

My response, which I can here summarize only briefly, is that both the theoretical and the empirical arguments against mental causation can be rebutted if we are careful in defining what we mean by “causation”. If we look at how causation is understood in the special sciences, this points towards a definition of causes as systematic difference-making factors for the resulting effects. (See, for instance, the interventionist theory of causation defended by Judea Pearl, 2000, James Woodward, 2003, and others.) Such a “difference-making” understanding of causation contrasts with a “production” understanding, which is typically assumed in epiphenomenalist arguments against mental causation (on the distinction, see Hall 2004). Roughly speaking, on a difference-making understanding, causal regularities are counterfactual regularities that remain in place when we control for confounding factors and which can be used for effective interventions in a system. An interest-rate increase by the central bank, for instance, is a difference-making cause of a reduction in inflation. Now, the most systematic difference-making causes of human actions are often at the intentional, psychological level, not at the sub-intentional, physical one; and this remains true even if – as we may very plausibly assume – there are underlying “producing causes” (as opposed to difference-making ones) at the physical level. Versions of this claim have been defended by several scholars (see, e.g., Woodward 2008, List and Menzies 2009, 2017, Raatikainen 2010, and Roskies 2012). It is our intentional, mental states that most robustly co-vary with the resulting actions, not their precise physical realizers in the brain, which are too fine-grained to qualify as difference-makers – a phenomenon that Peter Menzies and I called “realization-insensitivity”.

In Libet’s experiments, the neuronal readiness potentials measured prior to a subject’s formation of a conscious intention are, arguably, not difference-making causes of the actions, among other things because subjects can still abort an initially intended action after the neural activity has begun. The neuronal readiness potentials are best understood as belonging to the physical implementation mechanism of voluntary action. The intentional, psychological level remains a significant site of causal regularities, all the more so when we move away from the simple motor actions studied by Libet and consider more complex actions that involve sophisticated planning. And so, the idea of mental causation remains explanatorily indispensable as well.

5. What follows from this?

A skeptic might say: the present arguments only show that viewing people as intentional agents with alternative possibilities and mental causation is explanatorily useful: a convenient theoretical construct or fiction. But that doesn’t imply that this is what human beings are really like. Explanatory usefulness doesn’t imply reality, and the picture of human beings as choice-making agents conflicts with the more reductionistic picture given to us by the fundamental sciences. In reality, people are nothing but heaps of interacting particles.

There are two things to be said in response. First, science does not mandate adopting the reductionistic picture of human beings. To the contrary, the special sciences, from biology to the social sciences, support the alternative, non-reductionistic picture, and this picture is entirely compatible with the “physicalist” assumption that everything in the world is the result of underlying physical processes. Scientists recognize that even if everything is grounded in physical processes, many phenomena would be impossible to explain through the lens of fundamental physics alone. Higher-level explanations, such as those we find in fields ranging from biology to the social sciences, are indispensable. The theoretical point which tends to be missed by proponents of radically reductionistic approaches is that supervenience does not imply explanatory reducibility. (For further discussion, see List 2019b.)

Secondly, from a scientific perspective, our best guide to any questions about which entities or properties are real is given by our best scientific theories of the relevant domains. If we wish to find out whether electrons or neutrons are real, we must consult particle physics. Similarly, if we wish to find out whether the patterns of the climate are real, we must consult meteorology and climate science. This idea, defended by W. V. Quine (1977) and Arthur Fine (1984), is sometimes called the “naturalistic ontological attitude”. In line with it, I suggest that if we wish to find out whether human agency, choice, and mental causation are real, we must consult our best scientific theories of human behaviour, and as noted, these theories give a positive answer.

Putting these considerations together yields an indispensability argument for free will. The argument has two premises:

Premise 1: Our best explanations of human behaviour depict humans as choice-making agents: agents with goals and purposes, alternative possibilities to choose from, and causal control over their actions. This depiction is indispensable and compatible with the rest of science.

Premise 2: If postulating certain properties or entities is indispensable in our best explanations of a given phenomenon and compatible with the rest of science, then we are (at least provisionally) warranted in taking those properties or entities to be real.

If we accept the two premises, we arrive at the following conclusion:

Conclusion: We are (at least provisionally) warranted in taking intentional agency, alternative possibilities, and causal control over one’s actions to be real phenomena.

This, in a nutshell, is the core of the naturalistic case for free will. My argument is analogous to the standard naturalistic argument for realism about other properties or entities in science. Physicists are realists about particles, fields, and forces because postulating them is indispensable in our best physical theories. Biologists tend to be realists about cells, organisms, or eco-systems because postulating them is indispensable in the best theories within their domains. And psychologists, at least since the cognitive turn in their discipline, are realists about mental states and processes because postulating them is indispensable in psychological explanations. I suggest that the case for realism about intentional agency, alternative possibilities, and mental causation is

no different than that for other emergent, higher-level phenomena whose reality we seldom doubt: the weather, markets, economies, and so on.

It is important to note that the naturalistic case for any ontological commitments in the sciences is always provisional. New scientific developments might render certain postulated properties or entities dispensable even when they were previously considered indispensable. In the case of free will, if new scientific developments were to undermine the first premise of my argument – that our best explanations of human behaviour depict humans as choice-making agents – then I would no longer be able to uphold my conclusion. It is good scientific practice to acknowledge this point.

Moreover, the present kind of indispensability argument for realism about some property or entity is compelling only to the extent that our best explanations of the relevant phenomena exceed a certain quality threshold. If some entity or property occurs in our best explanation of a given phenomenon, but the explanation itself is very poor, then we cannot plausibly draw any ontological conclusions from this. I will here assume, however, that the required quality threshold is met in the case of many of the psychological and social-scientific explanations that depict humans as choice-making agents.

In the next section, I will say more about why the sort of agential indeterminism that underpins the present picture of agency is compatible with physical determinism.

6. Indeterminism as an emergent phenomenon

I have argued that realism about free will is justified because the picture of humans as agents with alternative possibilities and causal control over their actions is not just compatible with science but indispensable in some of our best explanations of human behaviour. Yet, one might wonder whether the picture of humans as indeterministic, choice-making agents is compatible with a worldview in which the laws of physics could, for all we know, be deterministic. Recall that determinism means that the state of the world at any point in time fully determines the future course of events. If the world is physically deterministic, then only one sequence of events will be physically possible, given the past. Everything that will happen in the future, including all human actions, will be inevitable consequences of the past. We would therefore have to be skeptical towards any theory that implicitly or explicitly postulates indeterminism in human agency.

I want to explain why this line of reasoning is mistaken. But before I do so, I need to make a few remarks about why this is relevant. One might think that quantum mechanics, one of our best physical theories, shows that the world is indeterministic. To give a simple example, when a photon, a light particle, hits a semi-transparent mirror with a very sensitive light detector attached, there is a 50% chance that the photon will be transmitted and a 50% chance that it will be reflected. Even the entire past history of the universe appears to be insufficient to determine which of these two outcomes will occur. If this is right, then the debate about whether there could be alternative possibilities in a deterministic universe is of no practical relevance, as our universe is indeterministic from the bottom up. However, this conclusion would be too quick. First of all, quantum mechanics is not the final word on physics. Notoriously, it has not yet been reconciled

with general relativity theory, which explains phenomena such as gravity, and that theory does not share the apparent indeterminism of quantum mechanics. The jury is still out on whether a future unified theory of physics will vindicate determinism or indeterminism. Secondly, the interpretation of quantum mechanics itself is controversial, and while some interpretations, such as the standard “Copenhagen” one, take it to imply indeterminism, others do not. Rival interpretations include ones according to which some hidden variables determine which trajectory the world is on. In the case of the photon, these hidden variables would have predetermined the photon’s path. This paper is not the place to discuss the interpretation of quantum mechanics. I simply want to note that the question of what physical determinism does not or does not entail is of more than hypothetical interest.

So, let me turn to the main question itself. Wouldn’t physical determinism rule out the kind of agential indeterminism to which, I have suggested, our theories of human behaviour are committed?

Suppose, for the sake of argument, the world is deterministic at the fundamental physical level. How, then, could there be any indeterminism in human agency? My answer begins with the observation that the physical level is just one among many different levels at which we may describe and explain the world, and other levels, such as the chemical, biological, psychological, and social ones, are no less important from a scientific perspective. Different such levels give us different windows into reality, and it would be a mistake to consider what we see from some of those windows as less real than what we see from others, especially when those windows correspond to well-confirmed scientific perspectives. When we are interested in what humans can and cannot do, the right level at which to ask this question is the level of the human and social sciences, not the fundamental physical one. This point should already be clear from what I have said so far.

But now comes a crucial point. Contrary to what is often assumed, the distinction between determinism and indeterminism cannot be drawn once and for all in a way that applies to all levels simultaneously. Rather, it is a level-specific distinction. The world may be deterministic at some levels and indeterministic at others – a point that may initially sound surprising.

To illustrate this point – as a “proof of concept” – let me introduce a toy model in which a system behaves deterministically at a micro-level and indeterministically at a macro-level (List 2014). Consider a system which, at each point in time, is in a particular state, and where that state evolves over time in accordance with certain laws governing the system. Let’s call the set of all possible momentary states in which the system could be its “state space”. A “history” of the system is a possible sequence of states across time. We can think of the system’s laws as constraints specifying which histories are possible and which not. For example, the possible histories could be as shown in Figure 1, reproduced from List (2019a). In this example, there are six time periods, labelled $t = 1, 2, 3, 4, 5, 6$. Little dots represent states of the system, and lines from bottom to top represent histories. We can think of the state in the bottom row as the system’s initial state, and we can think of the states along the upward-moving lines as the subsequent states. In this figure, all the possible histories are deterministic. That is to say: the initial state of each history fully determines all subsequent states; there is never any branching in any of the possible histories. We can interpret

the states in Figure 1 as micro-states of the system, for instance states that specify the complete configuration of all the physical particles, fields, and forces making up the system at the relevant time. Possible histories then represent the system's behaviour at a micro-level.

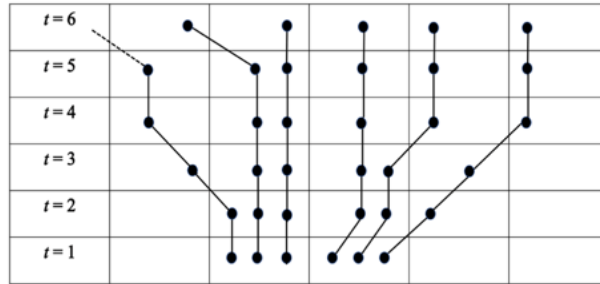


Figure 1: Deterministic lower-level histories

Now, let's suppose that we are interested in the system's behaviour at some macro-level, where the focus is not on particles, fields, and forces, but on certain macro-states. These "supervene" on the system's micro-states, but are more coarse-grained, in the sense that the same macro-state can be instantiated by different micro-states; they are "multiply realizable". An example of such a macro-state in physics is a system's temperature. Different configurations of molecules can have the same mean kinetic energy and thereby instantiate the same temperature. An example of a macro-state in psychology is a mental state such as desiring to eat chocolate and believing there is chocolate available in the kitchen. Plausibly, different neuronal configurations in the human brain could realize that same macro-state.

Formally, we can think of each macro-state as an equivalence class of micro-states, consisting of all its different possible "micro-realizers". In our example, suppose that whenever two or more different micro-states lie in the same cell of the rectangular grid in Figure 1, they instantiate the same macro-state. The relevant equivalence classes are thus given by the cells. While in this toy example there are no more than three possible micro-states for each macro-state, the real systems we study in the special sciences typically admit more complex forms of multiple realizability. In principle, each macro-state could have infinitely many possible micro-realizers, and it might be infeasible to describe what they all have in common from a micro-level perspective alone. Figure 2, also from List (2019a), shows what our toy system looks like at the macro-level. Thick dots represent macro-states, and thick lines from bottom to top represent macro-histories.

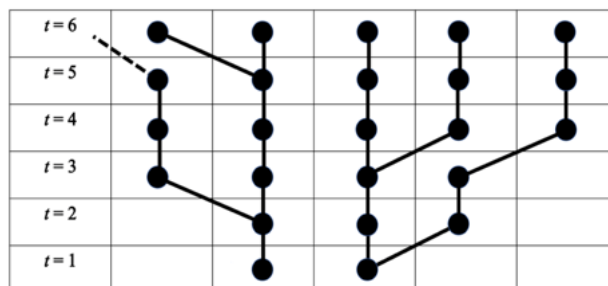


Figure 2: Indeterministic higher-level histories

It is easy to see that, unlike the micro-histories, the macro-histories are not deterministic here. Regardless of the system's macro-state at time $t = 1$, several sequences of subsequent macro-states are possible: the macro-histories exhibit branching. This illustrates that macro-level indeterminism, such as the indeterminism we find in the human and social sciences, can be an emergent byproduct of micro-level determinism. More technically, the property of determinism is not preserved under changes in the level of description, such as when we move from a lower, more fine-grained level to a higher, more coarse-grained one. Crucially, all of this is entirely consistent with the higher level supervening on the lower one.

Jeremy Butterfield (2012) has expressed the same point by saying that, in a system that admits multiple levels of description, the system's micro- and macro-dynamics need not "mesh". Furthermore, one can not only go from determinism at a lower level to indeterminism at a higher one, but the reverse is also possible (for a concrete illustration, see List 2019a, pp. 96–97). The bottom line is that indeterminism at the lower level is neither necessary, nor even sufficient, for indeterminism at the higher level. Related results were obtained by Jeffrey Yoshimi (2012) and, with a slightly different interpretational angle, by Charlotte Werndl (2009).

7. Some objections

I will now address a number of common objections.

7.1. *Isn't the emergence of indeterminism at a higher level merely epistemic?*

The most common objection to my analysis is that, if the fundamental laws of nature are deterministic, then the appearance of indeterminism at a higher level is merely "epistemic" – the result of our incomplete information about the system's micro-state – and so my defence of alternative possibilities fails. Even if the *macro-state* at time $t = 1$ is insufficient to determine the history of subsequent macro-states, the *micro-state* at time $t = 1$ would certainly fix all subsequent states, micro as well as macro. In consequence, what I have called "higher-level indeterminism" is an illusion due to our epistemic limitations.

However, this objection is mistaken. There are reasons for adopting an "ontic" and not merely "epistemic" interpretation of higher-level indeterminism – an interpretation which treats it as a real phenomenon. Let me sketch just a few of these reasons.

First of all, good scientific practice – in the spirit of the naturalistic ontological attitude – supports a form of pluralism about levels under which it is appropriate to take a realist attitude towards the properties at each level, provided they are explanatorily indispensable for some relevant special-science purposes and treating them as real is not incompatible with other, more fundamental commitments. Realism about higher-level indeterminism is arguably supported by this principle.

Secondly, the claim that the system's micro-state would be enough to fix all subsequent macro-states does not contradict macro-level indeterminism at all. It merely reasserts the already assumed fact that the system is deterministic at the micro-level. As an objection to macro-level

indeterminism it fails, because the definition of macro-level indeterminism does not – and should not – refer to the system’s micro-states. Indeed, the objection cannot even be expressed if, as is appropriate for macro-level descriptions, we refer only to macro-level facts. Macro-level indeterminism means that the system’s *macro-state* at a particular time does not determine the subsequent sequence of macro-states. This definition is unambiguously satisfied in Figure 2, and it is the right definition for analysing a system’s macro-level dynamics, also in line with the pluralistic case for considering each level on its own terms.

Finally, we cannot assume that there is always a most fundamental level, which can somehow be treated as the privileged level for distinguishing between determinism and indeterminism “simpliciter”. As Marcus Pivato and I have formally shown (2015), a scenario in which there is a bottomless hierarchy of levels, with determinism at even-numbered levels and indeterminism at odd-numbered ones, is entirely coherent, albeit hypothetical. In such a scenario, it would make no sense to speak of determinism or indeterminism “simpliciter”, or to tie the distinction to any particular privileged level; after all, there is no fundamental level here. The system’s indeterminism at odd-numbered levels is no more or less real than its determinism at even-numbered ones. This scenario supports the idea that the distinction between determinism and indeterminism is best understood as a level-specific distinction. And it fits nicely with the proposed ontic as opposed to epistemic interpretation of level-specific determinism or indeterminism, especially once we accept that different levels can be equally legitimate windows into reality, none of which is generally privileged over the others.

7.2. Isn’t agential indeterminism yet another form of randomness?

Another common objection is that, even if there is agential indeterminism, this only establishes a form of randomness or pseudo-randomness at the level of agency; but surely, this is not enough for free will.

I agree that free will requires more than randomness. First of all, however, we must not forget that agential indeterminism is only one of three requirements for free will; intentional agency and mental causation are needed too. But second and much more importantly, it would be a mistake to equate agential indeterminism with randomness. Randomness and indeterminism are not the same thing.

My analysis suggests that there are different kinds of indeterminism. Some are associated with randomness, for instance the kinds of indeterminism we find in quantum random generators or in statistical physics. In the human and social sciences, however, there is another kind of indeterminism, which is associated with option availability. In intentional explanations, we draw a crucial distinction between the options that an agent could possibly choose and those that the agent will actually choose (often for intelligible reasons). Agential indeterminism means that the set of possible options is non-singleton (meaning different courses of action are possible for the agent), not that the choice is random. And this is the kind of indeterminism required for free will, as well as the one supported by our theories of agency.

Generally, any definition of determinism is based on some underlying modal notion. Physical determinism is defined in terms of physical possibility. It is the thesis that each fully specified physical state of the world admits only one *physically possible* trajectory of future states. Physical indeterminism is the negation of this thesis. Biological determinism, if there is such a thing, would be defined in terms of biological possibility. It is the thesis that each fully specified biological state of a given system (which is more coarse-grained than any fully specified physical state) admits only one *biologically possible* trajectory of future states. Biological indeterminism is its negation. Agential determinism and indeterminism, finally, are defined in terms of agential possibility, in an analogous way. Agential possibility, in turn, is the notion of possibility used by our best theories of human agency, which I have suggested can be interpreted in terms of option availability as postulated by those theories. And while the theories represent human agents as making choices between different options – albeit perhaps not always fully rational choices – they do not represent human agents as mere randomizing devices. If we are committed naturalists, we should take this representation of human agents as choice-makers rather than randomizers at face value.

7.3 Are our best theories of human agency committed to real alternative possibilities or just to imagined ones?

I have adopted a realist account of agential possibility and suggested that agential possibility is the modal notion to which our best theories of human agency are committed. A critic might ask, however, whether for the sciences of human agency a purely *epistemic* notion of an agent's possibilities might suffice (a question posed, for instance, by Rosen 2020). We might interpret the possible options postulated by our theories of human decision-making as possibilities an agent imagines, rather than as real options. On such an epistemic interpretation, as I describe it in List (2019a, p. 102), “the ‘possible’ options would ... be those that an agent subjectively believes to be possible, even though in reality there is only one genuinely possible option in each situation – namely, the one that ends up being chosen”. This might be sufficient for explanatory purposes, without entailing any commitment to agential indeterminism.

I disagree with the suggestion that this is a good interpretation of what our theories of human action and decision-making are committed to. For a start, decision theorists – in fields like game theory and behavioural economics – do not normally think of the postulated options merely as options the agent *believes* he or she has, but as options that are *genuinely available* to the agent. Indeed, the use of concepts such as “consideration sets”, “focal options”, or “salient options” (which refer to *subsets* among the possible options to which a decision-maker gives particular attention) is evidence that decision theorists recognize that there is a further distinction to be drawn between options that are possible in some objective sense and options that have a certain subjective status in the decision-maker's awareness. I therefore maintain that the most literal and straightforward interpretation of the postulated options in decision theory (before we introduce notions such as consideration sets) treats them as real possibilities (from the external perspective of the decision theorist), not as imagined ones (from the internal perspective of a decision-maker).

To insist on adopting an epistemic interpretation of the modal notions of the human and social sciences while accepting an ontic interpretation of the modal notions of, say, chemistry or biology (which I think we normally do) would be to apply an unwarranted double standard. Consistently with the naturalistic ontological attitude, I propose that we should be realists about the modal notions of *all* of the special sciences, provided our commitment to them does not conflict with more fundamental commitments. And remember that my compatibility arguments show that what I say about agential possibility does not conflict with determinism at other levels, such as the physical one.

As I observe in my book, the epistemic interpretation of agential possibility “would amount to a kind of ‘error theory’ concerning the nature of human deliberation: agents would be systematically mistaken in thinking that they are faced with possible choices when in fact they never have more than one option” (List 2019a, p. 103). Such an error theory might be defensible if a realist interpretation were ruled out by other considerations, for instance if real alternative possibilities were ruled out by physical determinism. The error theory would then allow us to reconcile the *appearance* or *illusion* of choice with its lack of reality. But I have shown that alternative possibilities at the level of agency are not ruled out by physical determinism. And therefore, the error theory is not forced upon us, so we can accept the more natural realist interpretation of agential possibilities instead.

7.4. Isn't my account of free will vulnerable to the problem of present luck?

A further objection, raised by Al Mele (2020) and Gregg Caruso (2020), is that my account of free will suffers from the “problem of present luck” or, more fully, “the problem of present indeterministic luck”. Mele has argued that this problem poses a significant challenge for traditional libertarian accounts of free will, but not for standard compatibilist ones. He worries that my account may be vulnerable to it. Caruso, who is a hard incompatibilist, further worries that my account suffers from an additional problem of “constitutive luck”, to which even compatibilist accounts of free will are vulnerable. I set this additional problem aside in this paper; for my response to Caruso, see List (2020).

Let me begin by explaining the problem of “present luck”. Suppose we accept that free will requires a form of indeterminism in an agent’s choices. Each time an agent makes a choice, say between A and B, both options are genuinely open to him or her; each is a genuine possibility.

What this means is that the agent’s history up to the time in question has two possible continuations: one in which the agent does A and another in which he or she does B. Suppose, now, the agent chooses A. Does this qualify as a responsible choice? Clearly, A was not *necessitated* by the agent’s prior state; B would have been equally possible. Given the same prior psychological state, it was entirely possible for the agent to do B instead. This casts doubt on whether we can genuinely attribute the choice of A to the agent rather than to a random process.

As I put the worry in List (2019a, p. 108), “If several distinct courses of action are equally consistent with the agent’s full psychological state at the given time, then it is hard to see how the agent’s actual action could be any more attributable to the agent than any of the other possible alternatives would have been. Why should I count as the ‘author’ of my action if there was nothing in my psychological state that necessitated that action?” It seems, then, that the actual choice is to be attributed more to luck than to genuinely responsible agency. This is the problem of present luck. As the libertarian Robert Kane (1999, p. 217), who recognizes the problem, puts it: “If an action is *undetermined* at a time *t*, then its happening rather than not happening at *t* would be a matter of *chance* or *luck*, and so it could not be a *free* and *responsible* action.”

The critics are right that anyone who thinks that free will requires indeterminism at the time of choice must explain how an agent can be considered responsible for the resulting choices. By contrast, standard compatibilists do not face this problem. This is obvious in the case of compatibilists who think that free will does not require any alternative possibilities at all. For instance, the problem evidently does not affect those compatibilists who think that an agent is free if and only if he or she stands in an appropriate relation of authorship or endorsement to his or her actions, where this may be fully consistent with the resulting actions’ being determined by the agent’s character and motives. (Of course, such compatibilists may face a different kind of luck problem of their own, namely the problem of “constitutive luck”, as discussed by Caruso 2020, which is to explain how agents can be held responsible for the results of their character and motives the origins of which may have been beyond their control.) But even those compatibilists who accept an ability-to-do-otherwise requirement for free will but define it in conditional terms can avoid the “present luck” problem. Suppose I want and intend to do A, and go ahead and do A. If my ability to do otherwise simply consists in the fact that *if* I had wanted to do B instead, *then* I would have done so, the truth of this counterfactual in no way challenges the fact that I did A out of my own volition; indeed, in the actual world, in which I wanted and intended to do A, no other action would have been possible. The truth of the counterfactual does not compromise my responsibility at all.

The critics are also right that, because I take free will to require indeterminism at the level of agency, my account is more similar to traditional libertarianism from the perspective of the “present luck” problem than to standard compatibilism, and so I must confront the problem. What can I say in response?

In principle, I could respond to the “present luck” problem by drawing on and adapting one of the existing responses that have been proposed by libertarians. But my preferred response is to build on the observation that agential indeterminism is not the same as randomness. If someone has a choice between A and B, in which A and B are genuinely possible options, this does not mean that what the agent chooses is just a result of chance or randomness. As Wlodek Rabinowicz and I have argued in joint work, the genuine availability of more than one option does not preclude the intentional endorsement of one of the options by the agent, so that the chosen option stands out among the available alternatives (List and Rabinowicz 2014).

The choice of, say, A need not be attributed to luck just because B could have been chosen instead. According to my account, what renders the agent's choice a responsible one, over which he or she can be said have control, is the conjunction of three things:

- (i) According to our best explanatory theory, the agent intentionally endorses A and chooses on that basis.
- (ii) Some other choice, such as B, would have been possible too, though it may not have been equally intentionally endorsed.
- (iii) The agent's intention is a difference-making cause of his or her actual choice (here of A). That is, in the nearest (though perhaps not in more remote) possible worlds in which the agent has that intention, he or she makes the choice in question, and in the nearest possible worlds in which he or she does not have that intention, he or she does not make that choice.

According to my account, all three propositions can be simultaneously true, and so it would be a mistake to conclude that the agent's actual choice is just a matter of luck.

Although a version of this response might be available to proponents of traditional libertarianism too, my account arguably has an advantage. Traditional libertarians tend not to distinguish between different kinds of indeterminism. Rather, they take one kind of indeterminism to be fundamental – typically physical indeterminism – and argue that free will requires indeterminism of that kind. Now, if that kind of indeterminism is the same as the one that supposedly underlies ordinary chance or randomness, then it is easy to see why the worries about luck may arise. From the perspective of the modal notions involved, choice, on that picture, may look similar to randomness. By contrast, a key feature of my account is the distinction between agential and physical possibility, and between indeterminism at the agential level and indeterminism at the physical one. As already noted, I argue that agential indeterminism is a *sui generis* form of indeterminism that is due to option availability, not chance. And so, on my account, it is easier than on a traditional libertarian one to differentiate agential indeterminism from chance or luck.

7.5 Doesn't physicalism entail some form of reducibility of mental properties to physical ones?

My arguments for free will rest on a non-reductive physicalist view: a view according to which mental properties, while supervenient on physical properties, are non-identical to their physical realizers and causally and explanatorily significant. I reject the reductive physicalist claim that, for any mental property M, there exists a corresponding physical property P such that

- (i) necessarily, M and P are co-instantiated (“equivalence in satisfaction conditions”), and
- (ii) M and P can serve the same explanatory role (“substitutability for scientific explanatory purposes”).

I suggest that there are combinatorial reasons as to why clause (i) is not generally satisfiable and that, even if clause (i) were satisfied, there are further conceptual reasons as to why clause (ii) is

not generally satisfiable either. Critics have raised objections to both of these claims (Rosen 2020, Kaiserman and Kodsi 2021).

Regarding my combinatorial claim, a critic might object that if a mental property M supervenes on physical properties, then we may define a set consisting of all the possible physical configurations in which M can be instantiated; call its elements P_1, P_2, P_3 , and so on. Plausibly, there are infinitely many of them. It will then be true that M holds if and only if either P_1 holds or P_2 holds or P_3 holds, and so on. Could we then not simply identify M with the disjunction P_1 or P_2 or P_3 or ...? If we count this disjunction as a physical property, we have found a physical property that is necessarily co-instantiated with M , thereby satisfying clause (i).

My response is that the criterion for a property to count as physical is that the property is definable in the appropriate physical-level language, and I interpret this to mean that it can be defined using a finite (albeit possibly long) expression. All well-formed sentences of standard languages (natural or formal) are finite, even if there is no upper bound on the admissible length of any sentence. Now, there are strong combinatorial reasons as to why finite definability of a higher-level property in lower-level language is the exception rather than the rule.

Suppose specifically:

- (a) there are infinitely many possible states in which a physical system could be (which seems reasonable to assume), and
- (b) our physical-level language is countable (like all standard languages, from English to textbook logics).

Then, by assumption (a), there are uncountably many possible *sets* of states. By assumption (b), only countably many of them are describable using our given language, because the language permits only countably many expressions. This means that, in combinatorial terms, almost all sets of physical states (all but countably many) are not finitely definable in the given language. If a higher-level property such as M supervenes on physical properties, then there will certainly exist some set S consisting of all and only those possible physical states that count as instantiating M . But, as the present reasoning shows, it would be a highly exceptional case if that set S were also describable using our physical-level language. And so, finite definability of a higher-level property in lower-level terms is the exception rather than the rule.

This is an upgraded version of the classic multiple realizability argument against the reducibility of higher-level properties to physical ones (List 2019b), which originally goes back to Jerry Fodor (1974) and Hilary Putnam (1975). If this argument is correct, then I can reject the reductive physicalist claim that any mental property is identical to some physical property.

Suppose, however, against all odds, that we can find a physical property P with which the mental property M is necessarily co-instantiated. Would I then need to concede the reductive physicalist's claim that M and P are identical? I suggest in my book that this would be too quick, because M

and P may still differ in the explanatory roles they can serve. I write (on p. 69): “Because the intentional property features in semantic or logical relations, it can serve as an ingredient in intentional explanations of an agent’s behaviour, e.g., by rationalizing certain actions. The neurophysiological property, by contrast, can serve at best as an ingredient in causal explanations.”

A critic might worry that my analysis here either begs the question or conflates concepts and properties. As Gideon Rosen (2020) puts this point, “[t]he reductionist who identifies M and P will say that because M and P are identical, P has whatever logical and semantic properties M has”.

This is a fair objection, and I accept that I need to say more about my assumptions about property individuation. It is right that if I adopted an account of property individuation according to which any two necessarily co-instantiated properties are identical, then I would not be able to argue that M and P could differ in explanatory role even if they are necessarily co-instantiated. My response must therefore be to adopt a more fine-grained account of property individuation which allows two properties to differ in their explanatory role even in case they are necessarily co-instantiated.

While I am not here committing myself to any particular theory of property individuation (my methodology is to keep my analysis modular, so as to be able to plug in different such theories), I find it congenial to assume that properties are individuated in a suitable hyperintensional way. How exactly to do this is admittedly complicated, and I accept that there are some theoretical costs associated with individuating properties hyperintensionally. In any case, I want to emphasize that my arguments against the explanatory substitutability of M and P are subsidiary arguments. My main case against reductive physicalism rests on the combinatorial considerations summarized earlier.

8. Concluding remarks

I have characterized free will in terms of three properties – intentional agency, alternative possibilities, and mental causation – and suggested that people really have those properties. Specifically, I have offered a naturalistic indispensability argument for realism about all three properties and suggested that they are ontologically on a par with many other higher-level properties we readily postulate in the special sciences. The mistake in the various forms of free-will skepticism that have recently gained popularity lies in their failure to recognize free will as a higher-level phenomenon, and in their tendency to search for free will at a lower level than the one at which it can be realistically found.

My account of free will is libertarian in one respect and compatibilist in another. It is libertarian insofar as it accepts alternative possibilities as a requirement for free will and asserts that humans really have alternative possibilities. But it is compatibilist insofar as it renders this compatible with physical determinism; hence I have proposed the label “compatibilist libertarianism”.

While this may initially sound like a contradiction in terms, the consistency of my account is achieved with the help of the distinction between the agential level and the physical one. In

particular, the agential indeterminism required for free will does not conflict with physical determinism. Indeed, my analysis shows that the question of whether the fundamental laws of physics are deterministic or indeterministic is completely irrelevant to the question of whether there are alternative possibilities at the level of agency. The latter question is adjudicated, not by fundamental physics, but by our best theories of human behaviour, and these support the notion of choice between alternative possibilities as central.

There is still one important point on which incompatibilists about free will are right. Free will is not compatible with determinism *at the level of agency*. If our best theories of human behaviour were to give us a deterministic picture of human psychology, thereby refuting the sort of agential indeterminism I have defended, then this would also amount to a refutation of free will of the kind I have discussed. For the time being, however, we have good grounds for thinking that our best theories of human behaviour are not like this. They support agency, choice, and mental causation as real phenomena.

Acknowledgements

This paper builds on, and summarizes ideas from, my book (List 2019a). It draws on, and overlaps with, a series of blog posts on “The Naturalistic Case for Free Will” at the Brains Blog (August 2019), a debate with Gregg Caruso and Cory Clark (List 2020, Caruso 2020, and Clark 2020), and my exchange with Al Mele and Gideon Rosen at an APA symposium in January 2020. I have greatly benefitted from those scholars’ criticisms, as well as from the comments of Michael Esfeld, Christian Loew, Barry Loewer, Eddy Nahmias, and Kadri Vihvelin. My defence of free will as a higher-level phenomenon goes back to my 2011 working paper “Free will, determinism, and the possibility to do otherwise”, subsequently published as List (2014). I also build on my work with Peter Menzies on mental causation (List and Menzies 2009, 2017), with Marcus Pivato on determinism and chance (List and Pivato 2015), and with Wlodek Rabinowicz on alternative possibilities and intentional endorsement (List and Rabinowicz 2014). For brevity, I omit detailed literature references here; for full references, see my book. The two closest precursors to my account of free will are Anthony Kenny’s account in *Freewill and Responsibility* (1978) and Daniel Dennett’s in *Freedom Evolves* (2003). Both recognize the higher-level nature of free will, but neither develops the idea in the exact same way. Another precursor is Mark Siderits’s “paleo-compatibilism”. He argues that “the illusion of incompatibilism only arises when we illegitimately mix two distinct vocabularies, one concerned with persons, the other concerned with the parts to which persons are reducible” (Siderits 2008, p. 30). Finally, Sean Carroll (2016) defends a levelled understanding of free will too. Other fellow travellers in the quest to defend free will against scientific challenges include Carl Hoefer (2002), Jenann Ismael (2016), Alfred Mele (2014), Eddy Nahmias (2014), Adina Roskies (2006), and Kadri Vihvelin (2013).

References

Butterfield, J. (2012). “Laws, Causation and Dynamics at Different Levels.” *Interface Focus* 2(1): 101–114.

- Carroll, S. (2016). *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*. New York (Dutton).
- Caruso, G. (2020). “Why Free Will is Not Real.” *The Philosopher* 108(1): 67–71.
- Churchland, P. M. (1981). “Eliminative Materialism and the Propositional Attitudes.” *Journal of Philosophy* 78(2): 67–90.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA (MIT Press).
- Clark, C. (2020). “The Social Sciences Have No Use for Undetermined Free Will.” *The Philosopher* 108(1): 72–75.
- Coyne, J. (2014). “What Scientific Idea is Ready for Retirement?” Edge.org.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA (MIT Press).
- Dennett, D. (2003). *Freedom Evolves*. London (Penguin).
- Fine, A. (1984). “The Natural Ontological Attitude.” In *Scientific Realism*, edited by J. Leplin, 83–107. Berkeley (University of California Press).
- Fodor, J. (1974). “Special Sciences (or: The Disunity of Science as a Working Hypothesis).” *Synthese* 28(2): 97–115.
- Hall, N. (2004). “Two Concepts of Causation.” In *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. A. Paul, 225–276. Cambridge, MA (MIT Press).
- Harari, Y. N. (2016). “Yuval Noah Harari on Big Data, Google and the End of Free Will.” *Financial Times*, 26 August 2016.
- Harris, S. (2012). *Free Will*. New York (Simon and Schuster).
- Hofer, C. (2002). “Freedom from the Inside Out.” *Royal Institute of Philosophy Supplements* 50: 201–222.
- Ismael, J. T. (2016). *How Physics Makes Us Free*. New York (Oxford University Press).
- Kaiserman, A., and D. Kodsi (2021). “Review of *Why Free Will is Real*.” *Mind* 130(519): 987–996.
- Kane, R. (1999). “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism.” *Journal of Philosophy* 96(5): 217–240.
- Kenny, A. (1978). *Freewill and Responsibility*. London (Routledge).
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA (MIT Press).
- Libet, B., C. A. Gleason, E. W. Wright, and D. K. Pearl (1983). “Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act.” *Brain* 106: 623–642.
- List, C. (2014). “Free Will, Determinism, and the Possibility of Doing Otherwise.” *Noûs* 48(1): 156–178.
- List, C. (2019a). *Why Free Will is Real*. Cambridge/MA (Harvard University Press).
- List, C. (2019b). “Levels: descriptive, explanatory, and ontological.” *Noûs* 53(4): 852–883.
- List, C. (2020). “Free will: Real or Illusion.” *The Philosopher* 108(1): 61–66 (article) and 76–80 (replies to critics).
- List, C., and P. Menzies (2009). “Non-reductive Physicalism and the Limits of the Exclusion Principle.” *Journal of Philosophy* 106(9): 475–502.

- List, C., and P. Menzies (2017). “My Brain Made Me Do It: The Exclusion Argument against Free Will, and What’s Wrong with It.” In *Making a Difference: Essays on the Philosophy of Causation*, edited by H. Beebe, C. Hitchcock, and H. Price, 269–285. Oxford (Oxford University Press).
- List, C., and M. Pivato (2015). “Emergent Chance.” *Philosophical Review* 124(1): 119–152.
- List, C., and W. Rabinowicz (2014). “Two Intuitions about Free Will: Alternative Possibilities and Intentional Endorsement.” *Philosophical Perspectives* 28: 155–172.
- Mele, A. R. (2014). *Free: Why Science Hasn’t Disproved Free Will*. Oxford (Oxford University Press).
- Mele, A. R. (2020). “Free Will and Luck: Compatibilism versus Incompatibilism.” *The Monist* 103(3): 262–277.
- Nahmias, E. (2014). “Is Free Will an Illusion? Confronting Challenges from the Modern Mind Sciences.” In *Moral Psychology. Vol. 4, Freedom and Responsibility*, edited by W. Sinnott-Armstrong, 1–57. Cambridge, MA (MIT Press).
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York (Cambridge University Press).
- Putnam, H. (1975). “Philosophy and Our Mental Life.” In *Philosophical Papers. Vol. 2, Mind, Language and Reality*, 291–303. Cambridge: Cambridge University Press.
- Quine, W. V. (1977). *Ontological Relativity and Other Essays*. New York (Columbia University Press).
- Raatikainen, P. (2010). “Causation, Exclusion, and the Special Sciences.” *Erkenntnis* 73(3): 349–363.
- Rosen, G. (2020). “Comments on *Why Free Will is Real*.” Presented at an Author-meets-Critics Symposium at the 2020 Eastern APA conference, Philadelphia.
- Roskies, A. L. (2006). “Neuroscientific Challenges to Free Will and Responsibility.” *Trends in Cognitive Science* 10(9): 419–423.
- Roskies, A. L. (2012). “Don’t Panic: Self-Authorship without Obscure Metaphysics.” *Philosophical Perspectives* 26: 323–342.
- Siderits, M. (2008). “Paleo-Compatibilism and Buddhist Reductionism.” *Sophia* 47(1): 29–42.
- Steward, H. (2012). *A Metaphysics for Freedom*. Oxford (Oxford University Press).
- Van Inwagen, P. (1975). “The Incompatibility of Free Will and Determinism.” *Philosophical Studies* 27(3): 185–199.
- Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn’t Matter*. Oxford: Oxford University Press.
- Walter, H. (2001). *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*. Cambridge, MA: MIT Press.
- Werndl, C. (2009). “Are Deterministic Descriptions and Indeterministic Descriptions Observationally Equivalent?” *Studies in History and Philosophy of Science Part B* 40(3): 232–242.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. New York (Oxford University Press).

- Woodward, J. (2008). "Mental Causation and Neural Mechanisms." In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by J. Hohwy and J. Kallestrup, 218–262. Oxford: Oxford University Press.
- Yoshimi, J. (2012). "Supervenience, Dynamical Systems Theory, and Non-reductive Physicalism." *British Journal for the Philosophy of Science* 63(2): 373–398.