

Free Will: Real or Illusion

A Debate

With contributions by

Christian List, Gregg Caruso, and Cory Clark

This is a preprint of an exchange published in *The Philosopher*, Autumn 2019

The published version can be found at <https://www.thephilosopher1923.org>

© C. List, G. Caruso, and C. Clark

Free Will: Real or Illusion?

Christian List

If you are a regular reader of popular science writings, the chances are that you have come across articles or books arguing that free will is an illusion. Writers such as Sam Harris, Michael Gazzaniga, and Jerry Coyne, among others, have influentially claimed that the notion of free will – a person’s capacity to choose and control his or her own actions – is refuted by modern science. They argue that human actions are caused not by people’s conscious mental states, but by physical processes in their brains and bodies and the world at large, processes over which people have no control. If this is right, the idea of free will is a remnant of a superseded, unscientific worldview.

Although such free-will skepticism is not yet the established view in society, nor dominant in academic philosophy, it is gaining prominence, and if it became accepted as the truth, it would raise significant questions about the picture of humans as choice-making agents, and it would cast into doubt society’s practices of assigning responsibility to people for their actions. It seems problematic to hold a person responsible for something he or she did not freely do. Some skeptics are untroubled by these implications or even find them congenial. Given the problems with many of the world’s criminal justice systems, they think that we should focus less on responsibility and blame and more on rehabilitation and reintegration of offenders. But one can reject the traditional retributivist approach to punishment and support criminal justice reform without abandoning the idea of free will. There are better arguments for a more humane criminal justice system, based on considerations of dignity and restorative justice. Abandoning the idea of free will would have far-reaching implications for our human self-understanding as autonomous agents, independently of any implications for criminal justice. How could we genuinely deliberate about important decisions – such as which career to pursue, whom to marry, which ideology to endorse – if we didn’t view ourselves as free in making those decisions? The free-will skeptics would have to conclude that no such decisions ever involve real choices.

I think that free will can be defended against this growing skepticism, and in particular without denying science or redefining free will itself. My thesis, developed in my book *Why Free Will is Real*, is that, far from being undermined by a scientific worldview, the idea of free will is actually supported by it.

To sketch my argument, let me say more about the challenge. A reasonable starting point for any discussion of free will is the assumption that free will requires some or perhaps all of the following properties:

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

To be sure, not every proponent of free will thinks that all three properties are needed. Some argue that we could have free will without alternative possibilities to choose from; intentional endorsement of our actions may be enough. But I want to grant that free will, properly understood, requires all three properties: agency, choice, and control. So, I agree that free will would stand on shaky ground if one or more of these properties turned out to be illusory.

Now the skeptics argue: if we look at the world through a scientific lens, such as that of fundamental physics or perhaps neuroscience, then we find no support for agency, choice, and control. None of these properties are among the physical features of the world. At a physical level, there are only particles and molecules, fields and forces, and maybe neurons firing in the brain, all inexorably governed by the laws of physics. Agency, choice, and control are outdated fictions. And so, there is no free will.

Although the popular-science arguments against free will are seldom articulated fully rigorously, one can disambiguate different versions of the challenge, depending on whether the point of contention is intentional agency, alternative possibilities, or causal control. One kind of argument derives the unreality of free will from the claim that there is no room for intentional agency in a physical world. The idea that humans are agents with goals and intentions is an illusion. In reality, the argument goes, human beings are just bio-physical machines. I call this the “argument from radical materialism”. In academic philosophy, Patricia and Paul Churchland have developed such an “eliminativist” view about intentional agency.

A second kind of argument derives the unreality of free will from the claim that the laws of physics may be deterministic, meaning that the initial state of the universe, say at the Big Bang, pre-determined all subsequent events; so, there would be no room for alternative possibilities to choose from. You may think that you had a choice whether to read this article or not, but in reality, your decision was made for you by the world’s initial conditions. I call this the “argument from determinism”. An academic version of it is Peter van Inwagen’s so-called “consequence argument”.

A third kind of argument, finally, derives the unreality of free will from the claim that when any of us act, it is our brains that make us do it; we have no causal control over our actions. Our intentions, or any conscious mental states, are only by-products of subconscious physical processes, and those physical processes are the real causes of our actions. It is a mistake to attribute our actions to our intentions. I call this the “argument from epiphenomenalism”. Proponents of this argument often cite Benjamin Libet’s famous experimental study suggesting that voluntary movements, such as hand or finger movements, are caused by neuronal readiness potentials in the brain rather than any conscious decisions to perform those movements. More theoretically, Jaegwon Kim has argued that postulating mental causes of human actions, as distinct from physical causes, goes against some core principles of a “physicalist” worldview (such as the principle that every physical event has a physical cause or that a physical event cannot simultaneously have a physical cause and a non-physical one).

What can be said in response to these arguments? The first thing to say is that all three arguments take a reductionistic, physical perspective on human beings. They understand humans essentially as bio-physical systems: conglomerates of interacting particles. And when one takes this perspective, it is no surprise that one finds little support for agency, choice, and control. Physical descriptions of things, such as those given by fundamental physics or even by neuroscience, are unlikely to refer to intentional agency, alternative possibilities to choose from, and causal control over our actions. Physical descriptions refer instead to particles and molecules, fields and forces, and perhaps signal transmission between neurons in the brain. Crucially, however, this does not imply that agency, choice, and control are unreal. They are emergent, higher-level properties: properties that result from physical processes

but which cannot be described in physical terms alone. We can identify them only if we take a sufficiently non-reductionistic perspective on human beings, such as that of psychology or the human and social sciences. To be sure, taking that perspective involves no denial of fundamental physics; it just cannot be reduced to it.

To give an analogy: think of organisms, universities, or economic growth. If we look at the world solely through the lens of physics, such entities or properties will remain hidden from view. Biological and social features are absent from fundamental physical descriptions of the world. Yet it would not occur to us to deny their reality. From the fact that there are no universities in, say, particle physics, or that there is no economic growth in, say, chemistry, it doesn't follow that there are no universities or no economic growth at all. They are higher-level entities or properties: complex products of underlying physical processes but not reducible to them. What is true about those other higher-level entities or properties is equally true about agency, choice, and control. These, too, emerge from physical processes, but do not feature in fundamental physical descriptions of the world.

So far, I've suggested that some of the skeptical arguments against free will misfire because they take a perspective on human beings that is too reductionistic, seeing humans simply as bio-physical machines. But can one also say something positive in defence of agency, choice, and control?

Let's take a step back and ask how we usually determine whether something is or is not real. Why, for instance, do we think protons, electrons, and neutrons are real, while ghosts are not? The reason is that our best scientific theories support the existence of protons, electrons, and neutrons, while they don't support the existence of ghosts. A generally accepted method for establishing whether certain entities or properties are real is to consult our best scientific theories of the relevant domain. If postulating the entities or properties is indispensable from the perspective of those theories – that is, the theories couldn't explain their target phenomena without postulating those entities or properties – then this gives us good reasons to accept the entities or properties as real. If postulating them is not needed, then we have good reasons to doubt their existence. Postulating protons, electrons, and neutrons is indispensable for explaining many natural phenomena, and so they pass this “reality test”. Postulating ghosts is explanatorily useless, and so they fail the test. In philosophy, these ideas characterize the naturalistic approach to ontology defended by W. V. Quine and Arthur Fine, but they go back to William of Ockham.

It should be obvious that the present reality test can be applied not just to the physical sciences, but also to other domains. Organisms, universities, and economic growth, my earlier examples of higher-level entities or properties, easily pass the test. Biologists couldn't adequately explain the living world if they didn't recognize the existence of organisms, and social scientists couldn't explain social phenomena if they didn't recognize the existence of universities or economic growth. Of course, the test allows that new discoveries lead us to revise our views on what is or is not explanatorily needed. Our ontology is not dogmatically fixed but remains open for revision.

Now comes an important point: the test just described supports not only the reality of protons, electrons, and neutrons, or of organisms, universities, and economic growth, but also that of human agency, choice, and causal control. First consider intentional agency. Although the various human and

social sciences – from anthropology and psychology to sociology and economics – differ widely in their methods, one feature they have in common is that they understand humans as intentional agents: beings with goals and beliefs and a capacity for perception and thought, who act in intentional ways. By understanding people like this, we can make sense of their behaviour and explain many social phenomena. Why do people show up for appointments? Why do they vote the way they do? Why do they respond to incentives? All these questions can be answered if we understand people as intentional agents. If, instead, we took a reductionistic view and saw people as mere conglomerates of interacting particles, we would fail to understand their behaviour, missing important behavioural patterns while getting lost in an overload of micro-level information. We would miss the forest for the trees. At most, we might be able to explain some details of the neural functioning of the brain. But physics and neuroscience aren't remotely capable of explaining human behaviour in its breadth and flexibility. As Daniel Dennett has noted, we need to take an "intentional stance" towards humans, not a "physical stance", to make sense of their behaviour. Postulating intentional agency is explanatorily indispensable.

Next, let's turn to alternative possibilities. Once we explain people's behaviour by viewing them as agents, the idea that they make choices between different actions becomes indispensable too. When we answer questions such as "why does a particular person act in such-and-such a way?" or "what would he or she do in such-and-such a situation?", we suppose that the person is faced with some options, which are his or her possible choices, then deliberates about them, and eventually selects one option among the possible ones, in a way that makes some sense in light of the person's beliefs, goals, and preferences. This way of explaining behaviour would not get off the ground if we didn't suppose that people make choices. Even in economics, with its seemingly mechanistic picture of people as maximizers of utility, the notion of choice between different options is absolutely central. I call the idea that agents face choices among several possible options "agential indeterminism". Agential indeterminism does not mean that people's choices are random or arbitrary. After deliberation, they may well find some actions more rational than others.

However, even if you are persuaded that the idea of choice-making is indispensable in the human and social sciences, you might still worry that this idea conflicts with the notion that the physical universe is deterministic, a scenario that physicists have not yet ruled out (roughly speaking, classical physics implies determinism, while quantum physics can be interpreted one way or another). Recall that physical determinism means that the initial state of the universe was sufficient to pre-determine all subsequent events. Just as the precise date and time of the next solar eclipse in London is already fixed by the current configuration of the sun and other celestial bodies, so, the reasoning goes, all future human actions are fixed.

My response is that, surprisingly, indeterminism at the level of agency is compatible with determinism at the level of physics. The point is subtle and somewhat technical, but as I show in my book and in earlier work with Marcus Pivato, the distinction between determinism and indeterminism is a level-relative one. It is not meaningful to ask: is the world deterministic or indeterministic "as such"? The question becomes meaningful only once we specify the level of description relative to which we are asking it. A system can be deterministic at one level, such as the level of fundamental physics, and indeterministic at another, such as that of some special science, like biology, psychology, or sociology. When we move from a lower, microscopic level to a higher, macroscopic one, we may see a kind of

phase transition from deterministic to indeterministic dynamics. A system's micro- and macro-level dynamics need not "mesh", as Jeremy Butterfield puts it. So, the agential indeterminism we find in the human and social sciences does not conflict with physical determinism.

Finally, consider the idea of causal control over our actions. The skeptics say that when we act, it is our brains that make us do it; we are not in control. But those skeptics seldom define precisely what they mean by "causation" and instead employ the notion in relatively vague terms. I argue in my book that if we are careful in defining the idea of causation, the skeptical conclusions can be avoided. In the special sciences – from biology and medicine to the social sciences – causes are usually defined as systematic difference-making factors for the resulting effects, which remain in place even when we control for possible confounders. In line with this, it can be argued that the most systematic difference-making causes of human actions are often found at the intentional, psychological level, not at the sub-intentional, physical one. For instance, the causal model that attributes my choice of coffee over tea at breakfast to my preference for the former over the latter is likely to offer a better (more parsimonious, more robust) explanation of my behaviour than any rival causal model that searches for the cause at the level of neural firing patterns. Even in Libet's experiments of voluntary hand or finger movements, it is doubtful whether the neuronal readiness potentials in the brain qualify as difference-making causes of the subjects' actions. Subjects can still abort an initially intended action after the onset of the readiness potential. And in the case of more complex actions, involving deliberation and planning, it is implausible to search for causal regularities exclusively at the level of the physical brain. The intentional, mental level remains a site of causal relations. The bottom line is that, even if neuroscience has shed some light on the neural mechanisms that underpin human agency, the idea that people have some intentional control over their actions remains scientifically viable.

It's time to conclude. My case for free will rests on an indispensability argument. Its first premise is that the picture of human beings as choice-making agents is central to the human and social sciences; postulating intentional agency, alternative possibilities to choose from, and causal control over the resulting actions is explanatorily indispensable. My second premise is that if postulating certain entities or properties is explanatorily indispensable in a particular domain, then this gives us good reasons to accept those entities or properties as real. These two premises together imply that we have good reasons to accept agency, choice, and control as real. And since these are the core requirements for free will, we have good reasons to accept free will as real too. My argument is contingent on the nature of our best explanations in the human and social sciences. Should we ever arrive at a compelling reductionistic or deterministic theory of human behaviour – one that dispenses with the idea of humans as choice-making agents while still properly explaining their behaviour – then my first premise would no longer be true, and my conclusion would no longer follow. For the time being, however, I maintain that free will is no less real than other higher-level properties or entities, like organisms, universities, or economic growth, whose reality people seldom doubt.

Christian List (<http://personal.lse.ac.uk/LIST>) is Professor of Philosophy and Political Science at the London School of Economics. His book "Why Free Will is Real" was published by Harvard University Press in 2019. For a more detailed version of the ideas presented in this article, see also his series of blog posts, "The Naturalistic Case for Free Will", at: <http://philosophyofbrains.com/category/books/christian-list-why-free-will-is-real>

Why Free Will is Not Real: A Reply to Christian List

Gregg D. Caruso

Christian List presents a compelling case for the reality of *agency, choice, and control*. And while I agree with much of what he argues, I nevertheless think he's mistaken about the core issue: the existence of free will. I am willing to concede at the outset that agents engage in intentional goal-directed behavior, that our mental states and intentions are sometimes causally efficacious, and that we even make choices—though perhaps not in the sense List thinks. Yet it does not automatically follow from this that agents possess free will in the sense that has been of central philosophical and practical importance in the historic debate. This is because agency, choice, and control are potentially consistent with the kind of *free will skepticism* I defend, which maintains that who we are and what we do is ultimately the result of factors beyond our control (whether those be determinism, indeterminism, or luck) and because of this we are never morally responsible for our actions in the *basic desert* sense—the sense that would make us *truly deserving* of praise and blame, punishment and reward.

Before presenting my arguments against free will and basic desert moral responsibility, let me provide some background. Contemporary theories of free will can be divided into two general categories: those that endorse and those that are skeptical of the claim that human beings have free will. The former category includes *libertarian* and *compatibilist* accounts of free will, two general views that defend the claim that we have free will but disagree on its nature or its conditions. The main dividing line between the two pro-free will positions, libertarianism and compatibilism, is best understood in terms of the traditional problem of free will and determinism. *Determinism*, as it is commonly understood, is the thesis that every event or action, including human action, is the inevitable result of antecedent circumstances in accordance with the laws of nature. The traditional problem of free will and determinism therefore comes in trying to reconcile our intuitive sense of free will with the idea that our choices and actions may be causally determined by factors over which we have no ultimate control, that is, the past before we were born and the laws of nature.

Historically, libertarians and compatibilists have reacted to this problem in different ways. *Libertarians* acknowledge that if determinism is true, and all of our actions are causally determined by antecedent circumstances, we would lack free will and moral responsibility. Yet they further maintain that at least some of our choices and actions must be free in the sense that they are not causally determined. Libertarians therefore reject determinism and defend an indeterminist conception of free will in order to save what they maintain are necessary conditions for free will—the *ability to do otherwise* in exactly the same set of conditions and/or the idea that we remain, in some important sense, the *ultimate source/originator* of action. *Compatibilists*, on the other hand, set out to defend a conception of free will that can be reconciled with determinism. They hold that what is of utmost importance is not the absence of causal determination, but that our actions are voluntary, free from constraint and compulsion, and caused in the appropriate way. Different compatibilist accounts spell out requirements for free will differently but widely endorsed views single out responsiveness to reasons or connection of action to what one would reflectively endorse.

In the past, the leading form of skepticism was *hard determinism*: the view that determinism is true and incompatible with free will—either because it precludes the ability to do otherwise or because it is

inconsistent with one's being the ultimate source of action—hence, no free will. For hard determinists, libertarian free will is an impossibility because human actions are part of a fully deterministic world and compatibilism is operating in bad faith. My own reasons for favoring free will skepticism, however, do not depend upon the truth of determinism. Instead, I maintain that while determinism is incompatible with free will and basic desert moral responsibility, so *too* is the kind of indeterminacy required by the most plausible versions of libertarianism. In this way, my view resembles Derk Pereboom famous account since it maintains that we lack free will *whether or not* determinism is true.

Against compatibilism, the view that free will is compatible with the causal determination of our actions by natural factors beyond our control, I maintain that there is no relevant difference between this prospect and our actions being causally determined by manipulators. The basic idea behind this argument is that if an agent is causally determined to act in a particular way by other agents—say, a team of neuroscientists who manipulate the agent's brain—then the agent is intuitively not morally responsible for that action, even if they satisfy all the prominent compatibilist conditions on moral responsibility. Imagine, for example, that a team of neuroscientists manipulate Elizabeth by means of a device implanted in her brain. The neuroscientists manipulate her into stealing a book from the library by directly causing in her various wants, desires, beliefs, and intentions via a remote control. Even if we assume that Elizabeth's inner psychological states are causally efficacious, she does what she wants, she approves of her own behavior, etc., etc., intuitively she would not be morally responsible (in the basic desert sense) for stealing the book. The argument then continues by maintaining that since such manipulation cases resemble, in the relevant ways, agents in the normal deterministic case—i.e., in both cases the inner psychological states of agents would be causally determined by factors beyond their control—we must conclude that if agents fail to be morally responsible in cases of manipulation, they also fail to be morally responsible in the normal deterministic case.

Libertarian accounts, on the other hand, fail for a different reason. One variety of libertarianism posits indeterminism at the level of events and states and argues that such indeterminism is capable of preserving free will. (I assume this is the kind of libertarianism List has in mind when he talks about “agential indeterminism.”) Against such accounts, I advance the “luck” or “disappearing agent” objection, according to which agents are left unable to *settle* whether a decision/action occurs and hence cannot have the control in action required for basic desert moral responsibility. For an agent to be morally responsible for a decision in the sense relevant to free will, they must exercise a certain type and degree of control in making that decision. On this libertarian picture of free will, however, only events are causes, and free decisions are causally undetermined. Because all causation is event-causation, and the indeterminism needed for free will would have to occur at level of events and states, the agent can have no role in settling whether a decision occurs beyond the role it plays in agent-involving events. Hence, given the indeterminism required for a free decision on this account, the agent cannot *settle* whether the decision/action occurs and hence cannot have the control in action required for basic desert moral responsibility.

While there is, I contend, a variety of libertarianism that could, in theory, supply this sort of control, I argue that it cannot be reconciled with our best philosophical and scientific theories about the world

since it requires a questionable notion of causation where an agent, understood as a substance and not just a collection of events, is capable of causing various events without being causally determined to do so. Since these various views exhaust the options on which we have the sort of free will at issue, I conclude that free will skepticism is the only remaining position.

In addition to the arguments above, I also offer a second, independent argument for skepticism that maintains that regardless of the causal structure of the universe, free will and basic desert moral responsibility are incompatible with the pervasiveness of *luck*. At the heart of the argument is the following dilemma, which Neil Levy calls the *luck pincer*: either actions are subject to *present luck* (luck around the time of action), or they are subject to *constitutive luck* (luck that causes relevant properties of agents, such as their beliefs, desires, and intentions), or both. Either way, luck undermines basic desert moral responsibility since it undermines responsibility-level control.

Consider the significant role luck plays in our lives. First, there is the initial “lottery of life” or “luck of the draw,” over which we have no say. Whether we are born into poverty or affluence, war or peace, abusive or loving homes, is simply a matter of luck. It is also a matter of luck what natural gifts, talents, predispositions, and physical traits we are born with. Beyond this initial lottery of life, there is also the luck of what breaks one encounters during one’s period of self-formation, and what environmental influences are most salient to us. Combined, these matters of luck determine what Thomas Nagel famously calls *constitutive luck*—luck in who one is and what character traits and dispositions one has. Since our genes, parents, peers, and other environmental influences all contribute to making us who we are, and since we have no control over these, it seems that who we are is at least largely a matter of luck. And since how we act is partly a function of who we are, the existence of constitutive luck entails that what actions we perform depends on luck.

In addition to constitutive luck, there is also *present luck*—luck at or around the moment of a putatively free and morally responsible action or decision. Present luck includes any genuine indeterminism that may exist in the immediate causal chain leading to action, but it can also include an agent’s mood, what reasons happen to come to them, situational features of the environment, how aware they are of the morally significant features of their surroundings, and the like. It is a matter of present luck, for instance, whether our attention wanders at just the right/wrong moment or whether chance features of the environment prime our deliberation. I contend, following my friend Neil Levy, that the one-two punch of constitutive luck and present luck completely undermine basic-desert moral responsibility.

The problem with constitutive luck, for instance, is that an agent’s *endowments* (i.e., traits and dispositions) result from factors beyond the agent’s control. Now, I’m sure defenders of free will, will say that as long as an agent *takes responsibility* for her endowments, dispositions and values, over time they will *become* morally responsible for them (and perhaps even gain some control over them). The problem with this reply, however, is that the series of actions through which agents shape and modify their endowments, dispositions, and values are themselves significantly subject to luck—and, as Levy puts it: “We cannot undo the effects of luck with more luck.” Hence the very actions to which defenders of free will point, the actions whereby agents take responsibility for their endowments,

either *express* that endowment (when they are explained by constitutive luck) or reflect the agent's present luck, or both. Either way, responsibility is undermined.

How does List's emergentist account of free will fare against these arguments? Not very well, I contend. First, note that my own skeptical arguments do not appeal to reductionism, epiphenomenalism, or radical materialism in any way. In fact, the variety of free will skepticism I defend is able to acknowledge all the various distinctions compatibilists have made regarding the different degrees of control agents possess, while simultaneously denying that these differences provide the control in action necessary for basic desert moral responsibility. This is because acting in accordance with our own ends, or being responsive to reasons, does not mean those ends, or the ability to rationally act on them, are not themselves ultimately the result of factors beyond our control. For List to successfully defend the reality of free will, he would need to respond to the kinds of arguments advanced above—especially the luck pincer and the disappearing agent objection—otherwise he runs the risk of attacking a strawman and refuting only a weaker set of arguments.

Second, List himself acknowledges that determinism at the agential level would indeed be a real threat to free will, but he maintains that determinism at the physical level alone poses no such threat. Setting aside for the moment the question of whether deterministic laws at the *physical level* are consistent with the presence of indeterminism at the *agency level*, something I have serious doubts about but cannot explore here, List's account does require the rejection of causal determinism at the agential level. This is because List maintains that at the agential level causal determinism *is* incompatible with free will since it is incompatible with *genuine alternative possibilities* among which we can choose—i.e., the ability to do otherwise in exactly the same circumstances, holding the past and the laws of nature fixed. Hence, List appears to agree with me (though for different reasons) that traditional compatibilism fails to reconcile causal determinism and free will, at least not *tout court*, since at the level relevant to free will, List's account rejects determinism and appeals to a libertarian notion of agential indeterminism.

This brings us to the core issue: even if we assume for the moment that rejecting causal determinism and appealing to an emergentist notion of agential indeterminism is able to preserve genuine alternative possibilities, there is still an open question whether positing indeterminacy at the agential level preserves the relevant notion of free will any better than determinism does. I do not believe it does. To see why, consider the following dilemma.

If, on the one hand, we interpret List as saying that in cases of genuine choice, agents can intentionally endorse multiple, mutually exclusive courses of action, but it's a matter of genuine indeterminacy which they actually end up performing, then List's account ends up suffering from the very same problem other libertarian accounts do—i.e., agents would be left unable to *settle* which decision or action occurs. Imagine that Farah is torn between two courses of action: (a) attending an important meeting at work tomorrow morning or (b) calling in sick and spending the day with a friend who is in town for just one day. She has good reasons for doing (a) but she also has good reasons for doing (b). Whatever she ends up doing, we can say that it was “intentionally endorsed” and that she had “reasons for doing it,” since the action would be consistent with her values, aims, goals, intentions, and general wishes. If, however, indeterminacy is genuinely involved in the agential causal sequence, then it really

is a matter of luck which action she ends up performing. To make vivid the lack of control agents have over genuinely undetermined events, consider what would happen if God rolled back the relevant stretch of history to some point prior to an undetermined event and then allowed it to unfold once more. Since events would *not* unfold in the same way on the replay as they did the first time around, since these are genuinely undetermined, and nothing the agent does (or is) can ensure which undetermined possibility is realized, the outcome of this sequence (in this case the agent's decision) is a matter of luck. Such luck, skeptics argue, is responsibility-undermining.

If, on the other hand, we interpret List as saying that when an agent is confronted with two or more possible courses of action, their "intentions" and "reasons for action" settle which outcome occurs, then we face a different set of problems. Consider the case of Martin Luther, who provoked the Roman Catholic Church by criticizing some of its doctrines and practices. When he was summoned to the Diet of Worms in 1521 and was asked to renounce his criticism of the Church, he stood by his views and reportedly said, "Here I stand; I can do no other." If we interpret this as meaning that there is *only one* course of action that would count as "rational for Luther" or "psychologically salient," *given the person he is*, then this path would lead to a kind of psychological determinism, where given Luther's particular state of mind and constitutive character, he *would* lack the unconditional ability to do otherwise. This conclusion would be unacceptable for List, since it would conflict with his requirement for genuine alternative possibilities. Hence, on this interpretation, given the kind of person Martin Luther was—with his particular character, values, aims, goals, and reasons for action—the *only* thing he could have done in that situation (keeping everything fixed, including his personal psychology) was to reaffirm his criticism of the Church.

It seems, then, that whichever path List chooses, he ends up running into difficulties. The first path succumbs to the disappearing agent objection since the agent would lack the control in action needed to settle the outcome. The second path would lead to a kind of psychological determinism, where an agent's actions are determined by their constitutive character, values, aims, goals, and reasons for action. Furthermore, both paths would succumb to the luck pincer: since the agent's actions would either be subject to *present luck* (luck around the time of action), or they would be subject to *constitutive luck* (luck that causes relevant properties of agents, such as their beliefs, desires, and intentions), or both.

In conclusion, I agree with much of what List says, including his arguments against epiphenomenalism and eliminative materialism. Yet, when it comes to the core issue of free will, I think he's mistaken. Appealing to agential indeterminism does not solve the problem of free will or address the most powerful skeptical arguments. Instead, it leaves us wondering how indeterminacy at the agential level can preserve the control in action needed for basic desert moral responsibility and how it can avoid the luck pincer.

Gregg Caruso is Professor of Philosophy at SUNY Corning, Honorary Professor of Philosophy at Macquarie University, and Co-Director of the Justice Without Retribution Network (JWRN) at the University of Aberdeen School of Law. His research interests include free will, agency, and responsibility (both moral and legal), as well as philosophy of mind, cognitive science, neuroethics, moral psychology, criminal law, punishment, and public policy. His new book Rejecting Retributivism will be published later this year. <http://greggcaruso.com> / @GreggDCaruso

The Social Sciences Have No Use for Undetermined Free Will

Cory J. Clark

In the present issue, Christian List argues that free will can be upheld without 1) denying a scientific understanding of human behaviour, and without 2) redefining the concept of free will to the compatibilist definition (the ability to perform actions on the basis of rational deliberation in the absence of coercion). List argues that three features of free will – intentional agency, causal control, and alternative possibilities – are explanatorily indispensable for the human and social sciences, and that if certain properties are explanatorily indispensable, they are therefore to be thought of as real. In this short essay, I will first support List's views on the realities of intentional agency and causal control, but will then aim to explain why genuine alternative possibilities are incompatible with a scientific understanding of human behaviour. Finally, I will explain how and why the social sciences already have dispensed of undetermined agency as explanatorily relevant (and are much better for it). The notion of intentional human agency is only useful to the social sciences if it is causally determined by prior features in an agent's life (i.e., their genes and environments), and thus ceases to be free.

List argues that 1) humans are agents with goals and intentions, and that 2) these conscious mental states are also causally significant. He conceives of human agency and control as emergent, higher-level properties that result from physical processes but that cannot be described in physical terms alone. Although it does seem at least *possible* that conscious goals and intentions could be described in physical terms alone (humans may not be capable of this, but perhaps a more sophisticated computational system could thoroughly describe human behaviour without appealing to mental states), and although it is exceedingly difficult to explain precisely how something that feels non-physical such as conscious thought could influence the physical world, nonetheless, some aspects of human agency (such as conscious goals and intentions) *likely are* causal in the sense that they cannot be extracted from a causal sequence of events without interrupting or altering the sequence of events.

While I also defend the reality of human agency, my argument is somewhat different from List's. As documented by Roy Baumeister and colleagues, to the best of our scientific knowledge, conscious thought likely is needed for a variety of sophisticated human behaviours, such as complex verbal communication, including goal-directed teaching and learning and group negotiations, coordination, and planning. Although not all of the processes associated with these behaviours will be conscious all of the time, it does appear that these kinds of sophisticated social behaviours cannot happen without consciousness. Consciousness is also metabolically costly and often distracts from other tasks, so it seems implausible that humans would have evolved such a costly adaptation if it did not confer other benefits for guiding behaviour in useful ways. It is likely no coincidence that humans, the one species that most clearly displays evidence of conscious thought, have also managed to dominate, control, and enslave other species (the morality of this success is a separate issue). Although scientists and scholars more generally have yet to understand the relation between the experience of consciousness and the physical world, it seems likely that human conscious thought (including intentions, desires, and goals) is causally efficacious. In short,

these elements of List's defence of free will do not violate a modern scientific understanding of human behaviour.

List starts to stray from a scientific understanding of human behaviour in his discussion of "agential indeterminism", or the idea that a human at a given moment in time, holding constant everything in that person's life (and indeed everything in the universe) up until the moment a choice is made, could make more than one possible choice. He argues that this indeterminism does not mean that choices are random or arbitrary, but that after deliberation a person might decide that one action is more rational than another. But where does deliberation come from? And what factors influence whether a person decides that one action is more rational than another? It seems to me that one must defend one of two options here: either 1) such deliberations are caused by nothing, in which case they are random, or 2) they are caused by something (or likely many things), and are thus causally determined by prior things.

To use one of List's examples, let's imagine a person is deliberating whether to have coffee or tea. Something (or likely many things) previously existing in the universe must compel that deliberation in one direction or another (e.g., a genetically caused preference for one flavour profile, the culture the individual grew up in, the time of day influencing the desire for more or less caffeine, the price of each beverage, etc.). If we held constant all features in an individual's causal history that could rationally compel an individual to choose one option over the other (e.g., their genes, the culture they grew up in, the time of day), and the decision was still undetermined in that the individual could choose coffee or tea, then their decision is random or arbitrary—it is not caused by explanatory features of the situation.

If a social scientist wanted to explain why a person chose tea, they would not assert that it was "because they wanted tea" or "because they intended to choose tea" or even "because they have a goal to drink more tea". Or at least that would not be the end of the explanation. These explanations just push the question back. Why did they want it? Why did they intend it? Why did they have the goal? If we understand agency as part of a causally determined sequence of events, it is explanatorily useful to scientists. If, however, we understand agency as undetermined, it ceases to be useful and thus violates List's first premise, that postulating genuine alternative possibilities is explanatorily indispensable to the human and social sciences. I explain this below.

Before I do, let me clarify that social scientists do not need to chase causal histories all the way back to the Big Bang in order to defend a determinist position with regard to understanding the many causes of human behaviour. Invoking the Big Bang, I think, is an argumentative trick aimed at making determinism appear incomprehensible because humans are likely incapable of imagining how a physical event could have necessitated specific human behaviours – or likely even the existence of conscious humans. It is very hard to believe that once the Big Bang occurred, it was determined that I would type these words. It is much easier to believe that it was determined that I would type these words if you analyse events in my own life (I was invited to write this essay, my Ph.D. dissertation was on free will, my mom told me to avoid philosophy which inevitably unleashed my interest in the topic, etc.). Of course, there were countless causal factors along the

way, but nonetheless, each step along the way was caused by prior events, which eventually precede my existence (e.g., my dad also has a Ph.D. in psychology, and I share his genes).

These are the kinds of explanations social scientists concern themselves with. We can categorize these explanations as Nature vs. Nurture (as coined by Sir Francis Galton) or, to use more modern parlance, we can categorize these as genetic/biological and environmental explanations (and the interactions between the two). The more genetically caused the variation in a trait or behavioural tendency is (e.g., personality traits, educational attainment), the less environmentally caused the variation in that trait or behavioural tendency is. (There is also noise, but this can be thought of as primarily idiosyncratic, random, unexplained variance, and presumably not what free will proponents have in mind.) Explanatory factors are not freely chosen: genes are not chosen, many environments are not chosen (e.g., where you went to school, your parents and siblings), and those environments that *are* chosen according to some definition of the word “choice” (e.g., career path, romantic partners) are chosen because of some combination of genetic and environmental factors (and noise). There is no use in postulating some further entity beyond the combination of these factors, such as an undetermined free will that is at least partially immune to the causal history of an individual’s life.

Social scientists now distinguish between ultimate and more proximate levels of causation. The ultimate explains why humans evolved particular traits and tendencies (e.g., we have sex because sex facilitates reproduction and thus the spreading of one’s genes); the proximate explains more immediate reasons why humans do what they do (e.g., we have sex because it feels good to us). The causes List describes, such as intentions and choices, are even more proximate than that (e.g., we have sex because we choose to). As List contends, these causes are real in the sense that they often cannot be removed from the causal sequence without altering that sequence (if a human did not choose to have sex, they might not), but they are only of interest to social scientists insofar as they are caused by other pre-existing conditions and thus can be explained by those other pre-existing conditions.

Social scientists aim to explain variance in human behaviour by identifying causes that influence human behaviour in systematic ways. If human choices could not be explained by the complex array of other identified causes, and thus ceased to be explainable and predictable, they would be utterly irrelevant to social scientists. Social scientists would then only be able to explain such choices by appealing to mysterious idiosyncrasies of particular individuals, which do not qualify as scientific explanations for human behaviour (though they might qualify as noise).

Free will, then – if understood as indeterministic – fails List’s test of indispensability. Variance in human behaviour (such as deliberations and choices) that could never be explained or predicted by causal features of the situation (because it is undetermined) would not be helpful to social scientists, who aim to explain variance in human behaviour. If aspects of free will, including intentional agency and choice, are understood within a causal sequence of events, they likely *are* indispensable to science, but then they cease to fit the definition of free will that List defends.

It is also worth pointing out, briefly, that our conceptions of other humans as morally responsible agents may have evolved to facilitate moral judgment and punishment. We morally judge and punish others because judgment and punishment (and even the possibilities of judgment and punishment) tame the selfish behaviours of others. We might find better ways of shaping human behaviour to be less selfish and more cooperative, but new strategies likely should be empirically tested before we abandon the notion of human moral responsibility, regardless of whether humans have undetermined free will.

Many scholars have challenged the notion that humans have free will. In recent decades, these conversations have drifted out of the philosophical realm into the scientific one, challenging philosophers to deal with complicated causal stories about human behaviour that the scientific experts themselves are only just learning to grasp. These causes range from the movement of subatomic particles and chemical interactions (factors that feel far removed from the social worlds of humans), to more intermediary factors such as genes, bodily hormones, and neural firings, and finally to those factors that we can feel, observe, and explain without technology, such as family environment, social rejection, and the presence of a potential sexual partner. We may never be able to predict human behaviour with precision, but we have certainly managed to categorize various causes of human behaviour. Notions such as intention and choice are entirely valid scientific explanatory causes of human behaviour, but only when understood as part of a causal sequence of prior events and other causes. If one removes them from this causally determined sequence, they cease to be explanatorily useful.

Cory J. Clark is associate professor of social psychology at Durham University. Much of her research explores motivated cognition within the domains of morality and politics. <http://coryjclark.com> / @ImHardcory

Replies to Caruso and Clark

Christian List

First of all, I would like to thank Gregg Caruso and Cory Clark for their spirited, thoughtful, and fair-minded responses to my defence of free will. I am honoured by the attention they have given my work, and grateful for the opportunity to reply to them. As Caruso's and Clark's responses are rich and detailed, I cannot discuss all of their comments here, but I will focus on what I take to be their main points. I will discuss Caruso's and Clark's responses in turn.

Caruso begins by conceding that “agents engage in intentional goal-directed behavior, that our mental states and intentions are sometimes causally efficacious, and that we even make choices [albeit with qualifications]”. He thereby recognizes that there is “a compelling case for the reality of *agency*, *choice*, and *control*”. At the same time, he denies that we have free will. The fact that Caruso and I more or less agree on the presence of agency, choice, and control in humans but disagree on the reality of free will shows that he and I use different definitions of free will. For me, free will is primarily a metaphysical and psychological notion, and agency, choice, and control are jointly sufficient for it. For Caruso, free will is a more normative notion, and agency, choice, and control, as I understand them, are not enough. In particular, Caruso argues, “we are never morally responsible for our actions in the *basic desert* sense – the sense that would make us *truly deserving* of praise and blame, punishment and reward”. (For the record, I agree that agency, choice, and control by themselves are insufficient for moral responsibility, which I take to be a more demanding notion than free will alone. Moral responsibility, I assume, also requires the capacity for normative cognition, the absence of coercive conditions, and reasonable access to relevant information. But I think these additional requirements can in principle be satisfied. So, I don't share Caruso's scepticism about moral responsibility.)

Why does Caruso think that free will as I understand it (even in an agent who is capable of normative cognition, is not coerced, and has all relevant information) is insufficient for moral responsibility? The reason is this. Any defence of will free, he notes, will be *either* of a compatibilist kind that does not require any indeterminism at the level of agency, *or* of a libertarian kind that requires such indeterminism. In the former case, the agent's responsibility for any action is undermined by the fact that it was impossible for the agent to act otherwise. In the latter case, the agent's responsibility is undermined by the fact that what he or she does is the result of some indeterministic process and therefore not genuinely “settled” by the agent. Caruso calls this the “luck” or “disappearing agent” objection. In short, irrespective of whether one adopts a compatibilist account of free will or a libertarian one, the conditions for moral responsibility are not met. My account of free will is structurally on the libertarian side (albeit with the special twist that I argue that agential indeterminism is compatible with physical determinism), and so, on the face of it, my account seems vulnerable to Caruso's “luck” objection.

How can I defend moral responsibility against this objection? Caruso distinguishes between two forms of luck. One is “present luck”, which he defines as “luck around the time of action”. The other is “constitutive luck”, which he defines as “luck that causes relevant properties of agents, such as their beliefs, desires, and intentions”. He thinks both forms of luck undermine responsibility. Let me first consider “present luck”. Caruso's thought is that if an agent's choices are indeterministic, as my account asserts, then it is hard to see how the agent could be held genuinely responsible for them. This echoes a point made by Robert Kane in a 1999 article: “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.” I recognize this challenge and discuss it in a paper that I wrote

together with Wlodek Rabinowicz. My response is roughly this. We must recognize that indeterminism at the level of agency, as I defend it, is not the same as randomness. To say that I have a choice between A and B, and that both A and B are possibilities for me, is not the same as saying that whether I choose A or B is just a matter of chance or randomness. Rather, I argue that the following two claims can be true together:

- Two or more different actions are possible for the agent.
- The agent's chosen action is an intentionally endorsed one and thus stands out among the alternative possibilities.

If this is right, then my choice of A need not count as random or a matter of luck, just because choosing B would have been possible for me as well. What matters for responsibility is that I *chose* A rather than B, that I *intentionally endorsed* that choice, and that my intention was the *difference-making cause* of the resulting action. All of this can be true on my account, as I explain in my book, and so the conditions for responsibility can be met. The “present luck” objection misinterprets what I have called “agential indeterminism” as agential randomness. Agential indeterminism, however, is best interpreted as the availability of different options, not as a form of randomness.

What about “constitutive luck”, the second form of luck that Caruso takes to undermine responsibility? Imagine a bank robber who has freely robbed a bank. Let us suppose, in particular, that he intentionally chose to commit the robbery, that he could have acted otherwise, and that he was in control of what he did. Why might one still be reluctant to hold him responsible? A sceptic such as Caruso would say that the bank robber's intentional endorsement of his action is not enough for responsibility, but that we must look into the source of that intentional endorsement. Where do the bank robber's motives come from? Why does he have the goals and values that led him to commit the robbery? Suppose his current motives can be traced back to some earlier choice, and suppose, for the sake of argument, that he endorsed that earlier choice as well. Wouldn't that make us more inclined to hold him responsible for his present actions? The sceptic would respond that the earlier endorsement is not sufficient either, but that we must look into its source as well. Suppose, for instance, the bank robber had a difficult childhood and got entangled in a rough peer group as a teenager, and that all of this affected his subsequent life. In the absence of those conditions, he might not have become a bank robber. In short, his current actions are the result of constitutive luck, and he cannot be held responsible.

I think the principle underlying this line of reasoning is the following: an agent is not responsible for an action or mental state unless he or she is also responsible for the prior mental states that supported that action or mental state. And in accordance with the same principle, the agent is not responsible for those prior mental states unless they, in turn, can be traced back to some even earlier mental states for which he or she is responsible. As I note in my book, this would entail that people are responsible for far fewer things than commonly assumed, and possibly there would be very little room left for responsibility at all.

My response is that the present principle simply demands too much. Let me quote from my book: “The bank robber ... *can* be responsible for his action, even if he is not responsible for all of the prior steps within the agential history that put him on track to becoming a bank robber. Of course, some aspects of the bank robber's history may indeed reduce his overall culpability, and a judge may legitimately take evidence of that history into account in adjudicating the case. Personal hardship, a difficult social environment, and other factors may well be considered culpability-reducing. That said, if the bank robber committed his action intentionally, had alternative possibilities, and was in control

of what he did, then there is no reason not to treat this as a case of free will in the sense [that I have defended], even if we reach a more nuanced assessment of the bank robber's culpability in light of his prior history." In sum, my view is that responsibility in the everyday sense is less demanding than what might be called "ultimate responsibility", which is the sort of thing that Caruso appears to be looking for: responsibility not only for the action itself but also for the entire prior agential history.

This leads me to an important point of agreement with Caruso. Although I defend free will and do not share Caruso's far-reaching scepticism about responsibility, I agree that people's choices must always be carefully assessed in their context, and that contextual factors may sometimes be culpability-reducing. Although the bank robber in my example may well bear some responsibility for his present actions, it would be a mistake to think of this responsibility as "ultimate responsibility" in line with the overly demanding principle I have discussed. For this reason, Caruso and I agree that we should give up the sort of retributivism that is still quite entrenched in many criminal justice systems in the world, and that we should support an approach to criminal justice that is based on considerations of human dignity, restorative justice, and rehabilitation, not retribution.

Let me now turn to Clark's response. A key premise of my argument for the reality of free will has been that the picture of human beings as choice-making agents is central to the human and social sciences, and that postulating intentional agency, alternative possibilities to choose from, and causal control over the resulting actions is indispensable for explaining human behaviour. I have suggested that, insofar as those postulated properties are explanatorily indispensable and there is no conflict with the rest of science, realism about them is warranted. Clark agrees with my views on the reality of intentional agency and causal control, but disagrees with my claims about alternative possibilities. She claims that "genuine alternative possibilities are incompatible with a scientific understanding of human behaviour" and that "the social sciences already have dispensed of undetermined agency as explanatorily relevant (and are much better for it)".

To explain her point, Clark describes in detail how the social sciences seek to offer causal explanations of human behaviour, attributing behavioural patterns to factors ranging from "genes, bodily hormones, and neural firings" to "family environment, social rejection, and the presence of a potential sexual partner". Although causal explanations of human behaviour are often still in their infancy, it is undeniable that the social sciences have made significant progress in identifying causal factors influencing people's behaviour. As I acknowledge in my book, recent scientific findings suggest, for example, that the neurotransmitter dopamine affects impulsive behaviour, that our genes influence our political leanings, for instance whether we are liberal or conservative, and that risk attitudes are associated with brain patterns. It would seem that it is only a matter of time before a much broader range of human behaviours can be explained in similar causal terms – all the more so in the age of big data, where our evidence base for statistical and causal inferences is much greater than before.

Clark concludes from this that "[t]he notion of intentional human agency is only useful to the social sciences if it is causally determined by prior features in an agent's life (i.e., their genes and environments), and thus ceases to be free". And further: "If, however, we understand agency as undetermined, it ceases to be useful and thus violates List's first premise, that postulating genuine alternative possibilities is explanatorily indispensable to the human and social sciences." At first sight, this looks like a strong objection to my claims, but I now want to explain why it is mistaken.

The first thing to note is that explanations of human behaviour can fall into two categories: intentional explanations, and non-intentional or sub-intentional ones. Intentional explanations are those that

explain human behaviour – at least partly – in terms of people’s intentional mental states, for instance their beliefs and desires, goals and plans. Non-intentional or sub-intentional explanations are those that explain human behaviour without invoking any intentional mental states. For example, an explanation of someone’s aggression that refers to the person’s value system falls on the intentional side, while an explanation of his or her aggression that cites solely the underlying brain chemistry falls on the non-intentional or sub-intentional side. It should be clear that the notion of intentional agency is central to *intentional* explanations, not to non-intentional ones.

I readily admit that if, at some point in the future, the human and social sciences ended up replacing all intentional explanations with non-intentional ones, then this would constitute a profound challenge for the idea of free will. Crucially, however, this would be not so much because it would undercut my claims about alternative possibilities, but rather because it would go against the idea of intentional agency itself. We would then have to conclude that the picture of human beings as intentional agents is no longer needed in order to explain people’s behaviour.

However, Clark agrees with me that postulating intentional agency is explanatorily useful and, indeed, that it is a real phenomenon. So, she seems to be making the following two claims:

- Intentional agency is real.
- Intentional agents should be thought of as deterministic systems; they do not have alternative possibilities to choose from.

I can now explain where – as I see things – Clark’s objection goes wrong. As I have argued in my defence of free will, the present two claims – that intentional agency is real and that intentional agents do not have alternative possibilities – are in tension with one another. Once we view people as intentional agents in explanations of their behaviour, the claim that they face choices between different options must be accepted as a necessary presupposition. Intentional explanations typically proceed on the basis of a three-part hypothesis. That is, they explain human behaviour by supposing that (1) people face choices between two or more options, (2) they then deliberate about these options, where that process could range from fast and spontaneous to slow and rational, and (3) they finally select one option among the possible ones, in a way that is at least approximately intelligible (or even rational) in light of the person’s beliefs, goals, and preferences. To give just a few examples, we use this explanatory strategy to explain why people show up for work rather than stay at home (except in cases of illness or special circumstances), why they choose one consumption bundle rather than another, why and under what conditions they keep their promises, and so on. If we insisted that people never have more than one possible option, then this entire explanatory strategy would become unavailable from the outset. Notions such as deliberate, intentional choice make sense only if we assume the existence of different options to choose from. If there was only ever a single possible option for an agent, the notion of deliberation would become rather trivial, and we would not even be able to draw any explanatorily useful distinctions between rational and irrational choices.

As I put it in my book, “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. It is even unclear to what extent we would be warranted in ascribing agency to an entity in the first place unless we took that entity to have alternative possibilities.” Helen Steward has made a similar point in her defence of what she calls “agency incompatibilism”. On her account, agency itself requires some form of indeterminism.

At this point, Clark would presumably reiterate her claim that if we accept such an indeterministic understanding of agency, this undermines the very usefulness of intentional explanations: “If ... we understand agency as undetermined, it ceases to be useful.” But this claim, I think, rests on the same misinterpretation of agential indeterminism that we already encountered in the context of the “present luck” objection discussed earlier. The mistake is to interpret agential indeterminism as agential randomness, and therefore to equate it with unpredictability. Agential indeterminism, however, is better interpreted as the availability of different options, and is thus entirely compatible with the predictability of human behaviour.

To give a simple example, those who know me well will be able to predict that I practically never drink any alcoholic drinks. This is a fairly stable behavioural pattern. And yet, each time I go to a restaurant, it would be possible for me to order a glass of wine rather than a glass of water. The fact that I do not choose that option – and that this is even quite predictable in light of my preferences – in no way undermines the claim that I genuinely have the option, and that it would be possible for me to choose it. That I do not do so does not imply that it is impossible for me to do so.

In sum, the assumption that people have alternative possibilities to choose from is part and parcel of the picture of humans as choice-making agents. Insofar as that picture is explanatorily indispensable in the human and social sciences, so are alternative possibilities.

Let me end by thanking Gregg Caruso and Cory Clark once again for an intellectually rewarding exchange. I have greatly enjoyed the discussion with them.