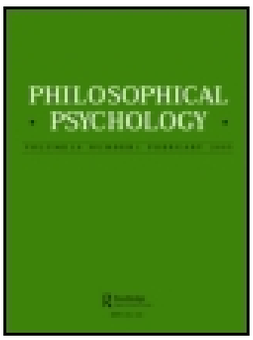

Philosophical Psychology

ISSN: 0951-5089 (Print) 1465-394X (Online) Journal homepage: http://www.tandfonline.com/loi/cphp20

"Local determination", even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn

Adina Roskies & Eddy Nahmias

To cite this article: Adina Roskies & Eddy Nahmias (2016): "Local determination", even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn, Philosophical Psychology, DOI: 10.1080/09515089.2016.1248286

To link to this article: http://dx.doi.org/10.1080/09515089.2016.1248286

Published online: 15 Nov 2016.

PHILOSOPHICAL PSYCHOLOGY, 2016
http://dx.doi.org/10.1080/09515089.2016.1248286

Routledge
Taylor & Francis Group

"Local determination", even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn

Adina Roskies[a1] and Eddy Nahmias[b1]

[a]Department of Philosophy, Dartmouth College, Hanover, USA; [b]Department of Philosophy, Georgia State University, Atlanta, USA

ABSTRACT

Marcelo Fischborn discusses the significance of neuroscience for debates about free will. Although he concedes that, to date, Libet-style experiments have failed to threaten "libertarian free will" (free will that requires indeterminism), he argues that, in principle, neuroscience and psychology could do so by supporting local determinism. We argue that, in principle, Libet-style experiments cannot succeed in disproving or even establishing serious doubt about libertarian free will. First, we contend that "local determination", as Fischborn outlines it, is not a coherent concept. Moreover, determinism is unlikely to be established by neuroscience in any form that should trouble compatibilists or libertarians—that is, anyone who thinks we might have free will. We conclude that, in principle, neuroscience will not be able undermine libertarian free will and explain why these conclusions support a coherent compatibilist notion of causal sourcehood.

ARTICLE HISTORY
Received 20 June 2016
Accepted 4 October 2016

KEYWORDS
Compatibilism; decision; determinism; free will; libertarian freedom; neuroscience

Introduction

Marcelo Fischborn discusses the significance of neuroscience for debates about free will. He explains why existing studies from neuroscience do not establish a type of determinism that threatens free will, a point with which we agree. He also raises important challenges for those of us who have questioned whether neuroscience or other mind sciences could possibly establish a type of determinism relevant to debates about free will. He argues that although Libet-style experiments have failed to threaten *libertarian free will*, that is, free will that requires indeterminism, in principle neuroscience and psychology could do so by supporting "less demanding forms of determinism" or "statements of local determination" (2016, p. 497).[1]

In what follows, we will argue that Fischborn is correct in his assessment of the extant Libet-style experiments, though we identify even more substantive problems for these studies. We then argue that he is wrong, that other experiments of this style could succeed where these fail. First, we doubt that "local determination", as Fischborn outlines it, is a coherent concept that can impact the free will debate. More specifically, local determination is unlikely to be established by neuroscience in any form that should trouble compatibilists or libertarians—that is, anyone who thinks we might have free will. In section 1, we discuss the failure of Libet's studies and others like them to undermine free will. In

CONTACT Adina Roskies adina.roskies@dartmouth.edu
[1] Authorship is equal.

section 2, we consider what kinds of determinism pose worries for the incompatibilist. In section 3, we argue that neuroscience cannot establish the relevant kind of determinism. We conclude that, in principle, neuroscience will not be able undermine libertarian free will. Finally, we explain why these conclusions help to support a coherent compatibilist notion of causal sourcehood.

1. Problems with Libet studies and their progeny

We agree with Fischborn's analysis of Benjamin Libet's studies and the more recent studies by John Dylan Haynes' group (see, e.g. Soon et al. 2008, 2013). Libet's original experiments purported to show that EEG traces of brain activity associated with a freely executed action, such as flexing one's wrist, preceded participants' awareness of deciding to move (Libet, Gleason, Wright, & Pearl, 1983). Averages of EEG traces time-locked to movement over many trials displayed a stereotypical signal associated with movement, called the *readiness potential* (RP), and the existence of the RP has been corroborated by many studies. Libet argued that the RP reflected the brain's unconscious decision and that conscious decisions were causally inefficacious because they occurred after the RP. While we take no issue with the existence of the RP, the original Libet experiments and interpretations of their significance were deeply flawed. For example, in Libet's experiments data were collected only in cases in which movement occurred, making it impossible to determine whether there were RP-like events that did not co-occur with movement. In fact, recent work by Schlegel and colleagues has shown that RPs can occur in the absence of motor activity, and thus that the presence of an RP does not unfailingly indicate the occurrence of a decision or intention to move (Schlegel et al., 2013). Further studies show that RPs occur during anticipation of a timed event or during a cognitive decision task (Alexander et al., 2016) and thus are not solely associated with movement. RPs even precede hypnotically induced actions which the subject may not be aware of making (Schlegel et al., 2015). All these results are consistent with the claim that the RP is a reflection of brain events of which we are unaware, but this is not enough to secure the claim that the brain events which the RPs reflect are unconscious *decisions*. Nor is the fact that we are unaware of these brain events evidence for our brains "making up our minds" before we do.

Indeed, it is unclear even that an RP corresponds to a particular brain event. Recent work by Schurger, Sitt, and Dehaene (2012) suggests that RPs could be a natural consequence of experiments that probe the events preceding action by averaging over spontaneous fluctuations in baseline activity over many trials time-locked to action initiation. If the decision to move is the result of decision processes that are described by an accumulator model, time-locking to movement and averaging small fluctuations would result in curves that appear as gradual exponential increases in neural activity prior to action, just like the RP. Independent work in neuroscience provides evidence for an accumulator model of decision-making in other areas. Thus, RPs are likely artifacts of the experimental design, given this popular neural model for decision-making.

Perhaps most importantly, the fact that we are unaware of the brain events underlying decisions tells us nothing about their role in cognition and action. After all, we are unaware of the brain events that constitute our conscious perceptual experience too, but these are paradigmatic cases of awareness! What made Libet's experiments seem compelling was that they seemed to provide evidence that participants had made their decisions without being conscious of having yet decided. However, the way in which Libet timed the awareness of decisions has been widely criticized as problematic (Banks & Pockett, 2007; Mele, 2010). Less often recognized is that the way in which Libet measured the time of awareness ensured that he was not probing the time of conscious decision, but rather another state—the awareness of our conscious decision, a meta-state which is causally dependent on that decision and thus must occur after the decision was made (Roskies, 2011). It is therefore unsurprising that this meta-state occurs late in the game, after decision-making neural processes are evident.

Since Libet's work, several other studies in the same vein also purport to show that decisions happen prior to the awareness of deciding. In one of these "Libet-style" experiments using neuroimaging, Soon and colleagues found predictive information in brain imaging data up to 10 s prior to the decision to press one of two buttons (Soon, Brass, Heinze, & Haynes, 2008). The results of these experiments

have been greatly exaggerated. To interpret the 57% accuracy found in these experiments (only 7% above chance) as evidence of *determination* of a decision is ludicrous. The data merely show that there exists some information about future choices in the brain, based on its previous activity. This is unremarkable. When choosing randomly, for instance, one retains a record of prior choices and future choices can be influenced by the nature of that record. If I have chosen one way three times in a row, my brain will have encoded that information, and I may be more likely to choose the other option the next time, so there will be a correlation between past and future choices (these studies use over 40 trials per participant). Thus, weak predictive power is unsurprising in these situations, and it indicates only that our decisions are sensitive to temporal factors such as prior history, not that they are determined (Lages, Boyle, & Jaworska, 2013).

However, if experiments of these sorts substantially increase predictive accuracy in the future, as Fischborn suggests they might, would they provide evidence for a type of determinism that challenges libertarian free will? We turn to this question now.

2. Kinds of determinism

Both Nahmias (2014) and Roskies (2006) argue that neuroscience cannot establish determinism. Nahmias (2014) argues that neuroscience and other cognitive sciences cannot establish determinism *as it is used in incompatibilist arguments*. Consider, for example, the two most influential arguments for the conclusion that free will is inconsistent with *universal* determinism—the Consequence Argument (e.g. van Inwagen, 1983) and the Manipulation Argument (e.g. Mele, 2013; Pereboom, 2001). The Manipulation Argument works by stipulating a powerful manipulator who has complete knowledge of the state of the deterministic universe, say, 30 years ago and of all the laws of nature who uses that knowledge to intervene on an agent (before or soon after his birth) to ensure that he does exactly what she wants 30 years later (while satisfying compatibilist conditions). The argument then challenges the compatibilist to either explain what difference there is between that agent and a "twin" agent whose un-manipulated life follows the same trajectory in a deterministic universe or to bite the bullet and say that such a manipulated agent has free will and is morally responsible for the manipulated action.

But if indeterministic events occur, even if they occur only *outside* the agent, the manipulator would be unable to ensure that the agent does what she wants—for instance, a few quantum events, magnified by non-linear dynamics, could radically alter the state of the universe in which the agent found himself 30 years down the line. Hence, to work at all, the Manipulation Argument seems to require universal determinism; local determinism—for instance, in the human brain—won't do the trick.[2]

Similarly, the Consequence Argument relies on universal determinism to reach its conclusion. The argument relies on a premise that no one has a choice about (or control over) the state of the universe in the distant past (or about the laws of nature), and (universal) determinism entails that the state of the universe in the distant past and these laws entail one's present actions. The argument does not work if there are indeterministic events anywhere that occur after the distant past and influence one's (much later) actions, since the entailment claim then fails. And the premise that no one has a choice about states in the past becomes much less intuitive if those past states are only a few moments ago, such that a more limited determinism might apply. For instance, if I chose to participate in a Libet-style experiment, then even if it could be shown that immediately prior brain events determine my actions, it would not be true that I had no choice about this fact, since I agreed to be in the experiment.

In sum, both of the prominent incompatibilist arguments rely on the postulation of *universal* determinism, as it is standardly defined in the free will debates (see, e.g. Hoefer, 2016; van Inwagen, 1983):

> **Universal determinism.** A universe is deterministic iff a complete statement P of the state of the universe at a time T, together with a complete statement of the laws L that govern that universe, entails a statement about any event E in that universe [at any time later than T]. That is, Necessarily((P&L) → E). (Some definitions leave out the bracketed phrase to indicate a time-symmetric form of determinism.)

Though some eschew defining determinism in terms of causation, there is another possible definition that might be used in modified versions of these incompatibilist arguments:

> **Universal causal determinism.** A universe is deterministic iff for every event E, there is some set S of prior events, such that, as governed by the laws L, S is a sufficient cause of E. That is, S causes E with probability 1. (Note that S is going to get larger and larger as the time between S and E increases and more background causes must be in place to ensure that E occurs.)

Now, it might be that incompatibilists could mount an argument that does not focus on universal determinism, but instead on a more localized version of determinism of the sort Fischborn describes. But it is unclear what such an argument would look like or whether it could be mounted without relying on premises and principles that rule out free will in an *indeterministic* universe as well, in which case it seems that local determination is irrelevant. For instance, Pereboom (2001), Strawson (1994), and others argue that free will is impossible whether determinism is true or false; it is not *determinism*, universal *or* local, that poses the problem, but rather the impossibility of agents' being originating causes or "ultimate sources" of their actions.

Kane (1996) and some other libertarians, such as Ekstrom (2000) and Balaguer (2012) certainly worry about local determinism in the brain, since their theories of free will require that indeterministic events occur in the brain (at the right time and place). However, they arrive at the conclusion that such events are required for free will only because they first develop or accept incompatibilist arguments using *universal* determinism (such as those described above). Only then do they posit indeterministic events at the time and place they think makes the most sense for free will, and they disagree among themselves where best to situate the indeterminism. They do *not* present incompatibilist arguments using local determination. Why should we accept the incompatibilism that motivates these libertarians unless we are considering the sort of *universal* determinism that neuroscience cannot establish, since only physics "covers" the entire universe (see Roskies, 2006)? Indeterminism anywhere, not just in the brain, falsifies the sort of determinism used in extant incompatibilist arguments.

As far as we know, there are no noteworthy arguments for the conclusion that local determination is incompatible with free will, and Fischborn offers no such argument.[3] Those who argue that neuroscience challenges free will sometimes talk as if local determination at the level of neural activity is what poses the problem, but that's not really what worries them. Instead, they focus on either the causal irrelevance of conscious mental states (i.e. epiphenomenalism)—for instance, because they think Libet's and others' studies show that such states occur too late to make a difference—or the causal irrelevance of non-physical mental states, because they think neuroscience establishes that dualism is false and our behaviour is caused only by physical states (or they think these two worries are the same) (see Nahmias, 2014).

But neither epiphenomenalism nor physicalism requires the truth of local (or universal) determinism. That is, indeterministic quantum events in the brain would not make a difference to whether epiphenomenalism or physicalism were true, though some have tried to posit that non-physical mental events somehow "use" quantum indeterminism as a place to make a difference (e.g. Popper and Eccles (1977). Conversely, local determination in the brain would not entail epiphenomenalism or physicalism. Conscious mental states can make a causal difference if the neural processes on which they supervene (or with which they are identical) make a causal difference, even if the causal relations are deterministic.[4]

So far, we have argued that local determination would not raise a challenge to free will without assuming universal determinism because either one has already bought into a libertarian view inspired by arguments that require universal determinism or local determination is not the real source of the challenge—for instance, because the real challenge is epiphenomenalism or physicalism or lack of causal sourcehood. Yet we might still wonder whether there is something intuitively threatening about local determination in the brain.[5] If so, we'd want to know if neuroscience could establish the truth of such local determination.

3. Could neuroscience establish local determination?

Fischborn's main point is that future neuroscience could establish a form of local determination that could challenge free will, but we will argue that he is mistaken. He suggests a scope-limited

determinism: for all events *of a certain type*, they are followed by another event, without exception. We first discuss the plausibility of such law-like relations. We then contend that either this type of local determination is not a threat to free will, because the event-types that it identifies are so limited as to leave open the possibility of agents' deciding otherwise, or they are so broad that neuroscience could never provide compelling evidence for their existence.

At first glance, Fischborn's claims seem plausible. Perhaps it is the way in which he couches his principles. For example, he refers to LD1 as a law.

> **LD1**. For any event *x*, and any subject *s*, if an *x* that is a pattern of neural activity of type *B* occurs in *s*'s brain, then *s* will decide to push a given button. (2016, p. 497)

Once we accept LD1 as a *law*, the rest of his argument seems more plausible. But why should we take it to be a law, without begging the question of whether such laws could be found? First, nothing in neuroscience would give us license to call LD1 a law rather than an empirical regularity. How do we know that for *any* event *x* of type *B* and *any* subject *s*, such a principle will hold? How do we generalize over subjects? What counts as the same pattern of neural activity across two different brains? For exceptionless laws, we look to physics, not to neuroscience. We doubt that neuroscience could give us principles like LD1 that can do the job.

Secondly, LD1 suggests that a pattern of neural activity *x* of type *B* causes *s* to decide to act with probability 1. Surely some brain pattern could cause one to act as if he had made a certain decision. For example, if we blasted *s*'s left motor cortex with enough electric current, *s*'s right index finger would move and push a button, if there was a button under that finger. But we would not call this finger movement the result of *deciding to push the button*. And there may in fact be normal patterns of brain activity that result in a right-sided button push, such as the patterns that constitute such a decision, or perhaps those that immediately precede reaching that decision threshold coupled with other activity that is sufficient to surpass it. Perhaps, once this pattern is instantiated the subject cannot choose to do otherwise. But this is not enough to show that he *could not have*, since earlier in the development of the pattern things could have been different. Merely passing a point of no return is not equivalent to lack of free will. I approach a traffic light in my car, and it turns yellow. I must decide whether to step on the gas or the brake. I decide. I step on the gas and accelerate before the intersection. Given how fast my car is going, I cannot now stop if the light turns red. That does not mean I did not freely decide to drive through the intersection rather than stop. Just because a decision may be irrevocable at a certain point in time does not mean it was not made freely (see Schultze-Kraft et al., 2016).

Furthermore, no one should be particularly worried about whether specific brain activity deterministically causes us to push a button, in part because it's a trivial behaviour (a "picking" not a "choosing") and in part because the determining neural activity (of type B) is probably *not* a "previous event whose occurrence was not within [the subject's] control" (the phrase Fischborn uses to set up his definition).[6] After all, if the subject just agreed to be in a neuroscience experiment that involves pushing a button whenever she feels like it (as in Libet's and Soon et al.'s paradigm), then her decision to agree to do so (or the neural activity underlying her decision) was both a significant cause for the existence of neural activity of type B, and it was under her control—unless, of course, this *earlier* decision was also not under her control. But now it looks like we need an argument to show that *all* of our decisions and behaviours are not under our control, or are such that we have no choice about them. Here we might go back to a standard incompatibilist argument using universal determinism, but if instead we're shifting our focus to local determination in the brain, we will want to see if neuroscience might show that claims like LD1 can be generalized to *all* decisions and behaviours. Fischborn suggests just this when he writes, "It should be clear that we could generate a potentially infinite number of statements of local determination like LD1" (2016, p. 497).

But it is not at all clear that we could generate infinitely many statements like LD1. That is, it is not clear that, even if future neuroscience might discover some deterministic "little laws" between trivial or automatic behaviours and immediately prior neural states, it will also discover deterministic causal relations between complex behaviours or decisions and *specific* prior neural activity. There are (at least) three reasons for doubting this sort of "neuro-determinism": multiple realizability, embodied and extended cognition, and complexity and chaos.

First, if multiple realizability is true, then there will not be LD statements that apply across all brains or even within the same brain over time (see, e.g. Laumann et al., 2015). Instead, many of the law-like relations, should they exist, will be among events at a higher level of description, and these events will be realized in a variety of neural processes. If there are any deterministic relations at the neural level, they would involve the mechanistic components (e.g. among neurons) that realize the higher level psychological events, such as perceptions and decisions, but these components will be organized in different ways within and between subjects as the realizers of the same psychological events. Hence, many of the causal generalizations involved in human decisions and actions would not be subserved by deterministic relations at the neural level.[7] Note that if such law-like causal relations exist at the psychological level, it is much less intuitively threatening to free will than determinism at the mechanistic level, because people interpret the latter to suggest that the causal work is being done in a way that leaves no causal work for our conscious deliberations and decision-making (see Nahmias, Coates, & Kvaran, 2007).

Moreover, if theories of extended or embodied cognition are on the right track (see, e.g. Clark, 2001, 2008; Varela, Thompson, & Rosch, 1992), there will be events and processes outside the brain, and perhaps outside the body, that are components of some of our mental states, including deliberations and decisions, and hence part of the set of sufficient causes for some of our behaviours. If so, neural activity alone will not be sufficient to cause some decisions or behaviours. Hence, local determination in the brain would not apply to these decisions and behaviours. For instance, suppose that some of the detailed information about the sentences of your paper which is stored in your computer cannot be held in your brain and that the display of this information on the computer screen is a crucial component (and not just a cause) of the information that you use as you make decisions about which words to write next. If so, then it is unclear how there could be law-like relations between neural states *alone* and your decisions.

Now let us consider the kind of neural pattern event-types that Fischborn invokes in LD1. Suppose that the event-type at issue is the firing of some connected neurons: whenever neurons A, B, and C fire, then neuron D fires. In principle, neuroscience could show that D, when monitored, always fires as a consequence of A, B, and C (but notice that this evidence would not demonstrate that it always *must*—we would only be observing a regularity, not proving that no other outcome could occur). However, nothing resembling a human *choice* or *decision* is encoded in such a small and restricted group of neurons, nor does the neuroscience of decision-making suggest that decisions are instantiated by necessary and sufficient conditions. We could not interpret such local determinism as the determination of a psychological state. When we take into account that each neuron in the brain makes, on average, approximately 10,000 connections to other neurons and is affected by as many, that there are multiple stages of processing, interactions with inhibitory interneurons, and crosstalk with other circuits for any actions we might take and decisions we might consider, we see that local determination would have to be fairly global to correspond to decisions, choices, actions, beliefs or other such complex mental processes or states. So even if deterministic relations can be pinpointed in the brain, they will occur in neural subsystems so impoverished that the neurons identified will not correspond to decisions or actions.

Conversely, any neural system rich enough to subserve the process of decision-making and its issuance in action will be too complex and interconnected to be governed by *local* determination. Alterations at any of multiple levels of processing could change the decisions; other thoughts, sensory inputs, or even noise could lead to a person's deciding or doing otherwise. There will be a requirement of global knowledge to establish determinism in a system, and global knowledge of that sort is unattainable (Roskies, 2006). Moreover, even if there were more widespread cases of local determination in the brain, some of the more promising libertarian theories are consistent with determinism in many cases of action, and thus do not depend upon indeterminism for all free actions. Kane's (1996) theory, for example, predicates freedom on indeterminism for self-forming actions (SFAs); these may occur infrequently and only under certain conditions. Subsequent actions are free in virtue of following from a character shaped by SFAs, yet could be entirely determined. Local determination for even a

large number of decisions would fail to be evidence for the claim that SFAs do not exist. Thus, even evidence of local determination would be insufficient to falsify a theory like Kane's.

Thus, the complexity of the human brain suggests that any complex decision will involve such a wide disjunction of neural processes that there will be no pattern of neural activity of a given *type* that can be picked out as the cause of that decision. Assuming physicalism and *universal* determinism, there will be a *token* set of neural activity *just prior* to the decision that ensures that it occurs. But that exact (token) set of neural activity will likely never happen again, nor will it ever happen in another agent. Throw in chaos theory (non-linear dynamics) and seemingly insignificant differences in any token pattern of neural activity might cause very significant differences in the ensuing decision or behaviour. Hence, the specificity of neural activity that determines later neural activity (such as the activity that realizes specific decisions) becomes increasingly fine-grained and the likelihood of there being law-like deterministic relations among complex neural states becomes vanishingly small.[8]

Fischborn suggests that finding enough local deterministic relations will decrease our confidence in the existence of Kane's SFAs, and thus in libertarian free will. Probably not. If one expects SFAs to be rare, then seeing some cases of local determination will not decrease credence in SFAs very much. Moreover, given what we know about stochastic activity in the brain, the likelihood of finding many local deterministic relations is small, and we suspect they would be quite local. And what should one conclude when one observes seemingly stochastic activity if one widens one's scope beyond that local network? Are the variations in activity merely effects of different but deterministic networks? Or do they reflect indeterministic activity? The answer matters greatly for how one updates one's credences, but it is unclear what reasons one has to choose one or the other option (Roskies, 2006). Thus, we maintain that to truly undermine libertarianism, you will need to find more than just evidence for instances of law-governed activity. You will need reasons to think that indeterministic events do not affect brain activity, and thus to think that global determinism is true. Mere evidence consistent with determinism will not suffice.[9]

4. Conclusions

Where does this leave things? We've argued that local determination of any sort that could be discovered in the brain will not provide evidence that bears significantly on the plausibility of libertarian theories of free will. We agree with Fischborn that Libet-style experiments fail to threaten free will, but for deep structural reasons and not just the idiosyncratic shortcomings that he discusses. If physics, contra the currently dominant interpretation of quantum physics, ends up discovering that the fundamental laws governing causal relations among the physical constituents of the universe are deterministic, then universal determinism will be true. Because humans are a part of the physical universe, that would mean that our decisions and behaviour would be causally determined. Incompatibilists worry about this possibility (along with some worrying that quantum indeterminism would not help). Compatibilists like ourselves need to address their arguments. We also need to address challenges from the modern mind sciences that suggest epiphenomenalism or related threats. But neither the incompatibilist arguments nor the scientific challenges arise because of the (limited, if any) deterministic relations among types of neural states that neuroscience might discover.

Seeing *why* neuroscience is unlikely to discover such deterministic relations at the level of human decision-making helps to illuminate why compatibilism makes sense. Our complex decisions cannot be causally explained (or predicted) by reference only to *types* of neural activity, much less by reference to *types* of states of the universe in the distant past (whatever that might mean). Instead, they often will be causally explained best by reference to our prior mental activity, which is realized in (token) neural activity (or perhaps neural activity plus states of the body or beyond). Our individual mental states as we deliberate and make decisions, realized in our complex and unique brains, thus serve as the *causal source* of many of our decisions and behaviours, even if our universe turns out to be deterministic.[10]

Notes

1. Fischborn uses the term "local determination" in his paper, and because we are discussing his arguments we follow his usage. However, as best as we can tell, he is talking about a scope-limited law-governed determinism, so we think that calling it "local determinism" would be more felicitous.
2. Deery and Nahmias (unpublished) argue that the manipulator could not pull off the trick even in a deterministic universe, since at the time she does her calculations she would not have access to events that are *outside* her past light cone—too far away to reach her even at the speed of light—but those unknowable events might be causally relevant to the agent's actions in 30 years, since they are *in* the past light cone of those actions.
3. Perhaps one could be developed if one was aiming for the conclusion Pereboom (2001) aims for: that free will is incompatible with one's actions being causally determined by factors beyond one's control. But Pereboom never defines exactly what he means by such causal determination, and it's not clear how the cases required for the Manipulation Argument work if universal determinism does not apply (see note 6 below).
4. Recognizing this point is crucial, we believe, in avoiding the misleading intuition that if our brains cause our decisions, *we* don't. Of course, we need to deal with the causal exclusion argument here (Kim, 2000), but that argument does *not* require deterministic causal relations at the lower level (though see Tse, 2013 for an argument that it does). See List and Menzies (in press) for a response to Kim's argument.
5. Nahmias, Shepard, and Reuter (2014) present experimental results suggesting that most people do *not* find free will to be undermined by the stipulation of perfect prediction of decisions based on preceding neural activity—which might suggest local determination in the brain.
6. It may be that Fischborn is gesturing here towards an argument that is similar to van Inwagen's Mind Argument (1983), something like: (1) No one has a choice about the neural activity that causes later decisions; (2) if no one has a choice about what causes X, then no one has a choice about X; (3) so, no one has a choice about their decisions. We think both of these premises are implausible.
7. By analogy, interactions among molecules might be deterministic (putting aside quantum effects for now), but any law-like causal relations among colliding billiard balls will be multiply realizable by different molecules and will be described in terms of the higher-level structures and interactions among the balls, not in terms of the laws that describe the molecules.
8. Note that Kane (1996) relies on non-linear dynamics to allow quantum indeterminism to "percolate up" to influence which among competing neural networks "wins out" to cause the outcome of a close-call decision (or SFA). But the chaos *without* the quantum indeterminism still gets us a lack of deterministic relations between specifiable types of neural activity and complex decisions or behaviours.
9. A reviewer helpfully alerted us to a manuscript by List and Pivato (2015) that provides a viable definition of local determinism, such that states within a spatio-temporal region L nomologically necessitate the states in a larger spatio-temporal region L*. We are not arguing against the theoretical coherence or possibility of local determinism of this sort. However, such a definition requires that L be a "closed" system that does not receive input from outside of that spatiotemporal region. The brain (and behaviourally relevant spatiotemporal subsets of it) is not a closed system, so the possibility of local determinism of this sort is not relevant to the question of free will.
10. There are a number of responses to incompatibilist worries about causal sourcehood. For example, Roskies (2012) explains how self-shaping can occur even in a deterministic world and uses interventionist theories of causation to show why it is legitimate to refer to an agent as the causal source of his or her own character. Deery and Nahmias (2016) also use causal interventionism to provide a compatibilist account of causal sourcehood. They apply this account to the Manipulation Argument to elucidate a principled difference, relevant to free will and moral responsibility, between a manipulated agent and a (merely) determined agent. The manipulated agent is not the causal source of his actions because there is another source (the manipulator) that better predicts and explains his decision. The same is not true for the determined agent.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A. L., Wheatley, T., & Tse, P. U. (2016). Readiness potentials driven by non-motoric processes. *Consciousness and Cognition, 39*, 38–47.

Balaguer, M. (2012). *Free will as an open scientific question*. Cambridge, MA: MIT Press.

Banks, W. P., & Pockett, S. (2007). Benjamin Libet's work on the neuroscience of free will. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 657–670). Malden, MA: Blackwell.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. New York, NY: Oxford University Press.

Clark, A. (2008). *Supersizing the mind*. New York, NY: Oxford University Press.
Deery, O., & Nahmias, E. (2016). Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood. *Philosophical Studies*. Advance online publication. doi:10.1007/s11098-016-0754-8
Deery, O., & Nahmias, E. (unpublished). The light-cone limit.
Ekstrom, L. (2000). *Free will: A philosophical study*. Boulder, CO: Westview Press.
Fischborn, M. (2016). Libet-style experiments, neuroscience, and libertarian free will. *Philosophical Psychology, 29*, 494–502.
Hoefer, C. (2016). Causal determinism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring, 2016 ed.). Retrieved from http://plato.stanford.edu/determinism-causal/
Kane, R. (1996). *The significance of free will*. New York, NY: Oxford University Press.
Kim, J. (2000). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge, MA: MIT Press.
Lages, M., Boyle, S. C., & Jaworska, K. (2013). Flipping a coin in your head without monitoring outcomes? Comments on predicting free choices and a demo program. *Decision Neuroscience, 4*, 925.
Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M. Y., … Peterson, S. E. (2015). Functional system and areal organization of a highly sampled individual human brain. *Neuron, 87*, 657–670.
Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain, 106*, 623–642.
List, C., & Menzies, P. (in press). My brain made me do it: The exclusion argument against free will, and what's wrong with it. In H. Beebee, C. Hitchcock, & H. Price (Eds.), *Making a difference: Essays in honor of Peter Menzies*. Oxford: Oxford University Press.
List, C., & Pivato, M. (2015). Dynamic and stochastic systems as a framework for metaphysics and the philosophy of science. Retrieved from https://arxiv.org/abs/1508.04195
Mele, A. R. (2010). *Effective intentions: The power of conscious will*. New York, NY: Oxford University Press.
Mele, A. (2013). Manipulation, moral responsibility and bullet biting. *The Journal of Ethics, 17*, 167–184.
Nahmias, E. (2014). Is free will an illusion? Confronting challenges from the modern mind sciences. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 4: Free will and moral responsibility* (pp. 1–25). Cambridge, MA: MIT Press.
Nahmias, E., Shepard, J., & Reuter, S. (2014). It's OK if "my brain made me do it": People's intuitions about free will and neuroscientific prediction. *Cognition, 133*, 502–516.
Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy, 31*, 214–242.
Pereboom, D. (2001). *Living without free will*. New York, NY: Cambridge University Press.
Popper, K. R., & Eccles, J. C. (1977). *The self and its brain*. New York, NY: Springer-Verlag.
Roskies, A. L. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences, 10*, 419–423.
Roskies, A. L. (2011). Why Libet's studies don't pose a threat to free will. In W. Sinnott-Armstrong (Ed.), *Conscious will and responsibility* (pp. 11–22). New York, NY: Oxford University Press.
Roskies, A. L. (2012). Don't panic: Self-authorship without obscure metaphysics. *Philosophical Perspectives, 26*, 323–342.
Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A., Tse, P. U., & Wheatley, T. (2013). Barking up the wrong free: Readiness potentials reflect processes independent of conscious will. *Experimental Brain Research, 229*, 329–335.
Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A. L., Tse, P., & Wheatley, T. (2015). Hypnotizing Libet: Readiness potentials with non-conscious volition. *Consciousness and Cognition, 33*, 196–203.
Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., … Haynes, J. D. (2016). The point of no return in vetoing self-initiated movements. *PNAS, 113*, 1080–1085.
Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences, 109*, E2904–E2913.
Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*, 543–545.
Soon, C. S., He, A. H., Bode, S., & Haynes, J.-D. (2013). Predicting free choices for abstract intentions. *Proc. Natl. Acad. Sci. U.S.A., 110*, 6217–6222.
Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies, 75*, 5–24.
Tse, P. U. (2013). *The neural basis of free will*. Cambridge, MA: MIT Press.
van Inwagen, P. (1983). *An essay on free will* (Revised ed.). New York, NY: Oxford University Press.
Varela, F. J., Thompson, E. T., & Rosch, E. (1992). *The embodied mind: Cognitive science and human experience* (New ed.). Cambridge, MA: MIT Press.