Philosophical Studies
https://doi.org/10.1007/s11098-022-01882-7

ORIGINAL RESEARCH

Why the manipulation argument fails: determinism does not entail perfect prediction

Oisin Deery[1,2] · Eddy Nahmias[3]

Accepted: 31 August 2022

Abstract

Determinism is frequently understood as implying the possibility of perfect prediction. This possibility then functions as an assumption in the Manipulation Argument for the incompatibility of free will and determinism. Yet this assumption is mistaken. As a result, arguments that rely on it fail to show that determinism would rule out human free will. We explain why determinism does not imply the possibility of perfect prediction in any world with laws of nature like ours, since it would be impossible for an agent to predict with certainty any future event that is causally influenced by events outside her own backward light cone yet inside the backward light cone of the future event. This is the *light-cone limit* and it undermines the Manipulation Argument or limits what this argument can tell us about the relevance of determinism to free will. We also respond to objections that the light-cone limit is irrelevant to the Manipulation Argument.

Oisin Deery and Eddy Nahmias: Authorship is equal.

✉ Oisin Deery
oisin@oisindeery.com

✉ Eddy Nahmias
enahmias@gsu.edu

[1] Department of Philosophy, Macquarie University, Levels 6 and 7, 25B Wally's Walk, 2109 Sydney, NSW, Australia

[2] Department of Philosophy, York University, S448 Ross Building, 4700 Keele Street, M3J 1P3 Toronto, ON, Canada

[3] Department of Philosophy, Georgia State University, 25 Park Place, Suite 1600, 30303 Atlanta, GA, USA

Published online: 10 November 2022

1 Strong vs. weak incompatibilism

Incompatibilism claims that it is impossible for any agent to have free will in a world governed by deterministic laws. In responding to specific arguments for incompatibilism, some have noted that these arguments work only if extra stipulations are added, or certain deterministic universes are not under consideration. For instance, Joseph Campbell (2007, 2008) has argued that the Consequence Argument for incompatibilism fails if the determined agents under consideration are *eternal*, such that they have no remote past over which they lack control. Alfred Mele, in response to his own Zygote Argument (discussed below), points out that the "original design" thought experiment on which it depends requires that the laws of nature be necessitarian, not *Humean* (2006: 194–5), and Helen Beebee & Mele (2002) argue that other incompatibilist arguments require that the laws also be *non-Humean*. These moves suggest that the incompatibilist arguments fail to show that determinism *alone* rules out free will. Instead, they can show, at most, that determinism supplemented by some extra condition, *X*, might rule out free will. Or, in terms of possible worlds, such arguments cannot establish the following *strong* incompatibilist thesis: For *any* possible world in which determinism is true, no agent has free will. Instead, they can only establish a *weaker* thesis: For any world in which determinism *and X* are true, no one has free will.[1]

These examples may offer little solace to compatibilists hoping to secure human free will, since the conditions we mentioned above are unlikely to be relevant to *humans*. After all, humans are not eternal agents and do have a remote past. And Humeanism—while possibly true—is not the favored interpretation of laws of nature.[2]

In this paper, we consider another influential incompatibilist argument, the Manipulation Argument, and we demonstrate that it does not show that determinism alone is incompatible with free will. Rather, the argument might show, at best, that determinism rules out free will only if determinism *also* permits perfect prediction (and subsequent manipulation) of events in the future. Perfect prediction is required for the cases stipulated by the Manipulation Argument to be possible. But such prediction requires, in turn, a world that, unlike ours, is *not* subject to what we call a *light-cone limit* (see Sect. 3). Thus, the Manipulation Argument requires not only the assumption of determinism but also that any deterministic world under consideration not be subject to a light-cone limit. So, the most incompatibilists can conclude from the Manipulation Argument is that some deterministic worlds with different laws of nature than our own contain no free agents.

Unlike the other *X*-conditions mentioned above, which do not apply to humans, a light-cone limit *does* apply in our world. Given the most plausible theories about how our laws of nature might be deterministic, these laws rule out the sort of prediction required for the Manipulation Argument. As a result, the Manipulation Argument

[1] See Warfield (2000) and Mickleson (2019), who raise related concerns about how to understand incompatibilism in stronger and weaker forms.

[2] Even if our laws *were* Humean, Beebee & Mele (2002) argue that the problem of luck would still raise difficulties for human free will.

fails to show that human free will would be threatened if the actual world is governed by deterministic laws. Thus, the Manipulation Argument fails as an argument *either* for the strong incompatibilist thesis *or* for the conclusion that determinism precludes human free will.[3]

2 The manipulation argument

The Manipulation Argument aims to establish the incompatibility of determinism and free will by showing that actions causally determined by a manipulator are not free and there is no relevant difference between such manipulation and determinism (Pereboom, 2001, 2014; Mele, 2006, 2013, 2019). Here, we will follow Derk Pereboom (2014: 2) in understanding the relevant notion of free will (and acting freely) to be the type of control in acting that would make an agent morally responsible for actions in the "basic desert" sense. Usually, determinism is characterized in one of two ways. On the *causal* thesis of determinism, for every event, *E*, the laws of nature and some set of events that occurred prior to *E* are such that these events cause *E* to occur with probability 1 (Ismael, 2013). On the *entailment* thesis, a conjunction of a complete statement of the (non-relational) facts of the world at a time with a complete statement of the laws entails all other (non-relational) facts about the world at other times (van Inwagen, 1983).[4]

Neither formulation is explicitly stated in terms of prediction. Yet both formulations are typically taken to *imply* something about prediction (Laplace 1814/1951). For instance, it is often assumed that in a deterministic universe, a suitably powerful being who knows all of the events occurring at one time (or all the facts of the world at that time) and who knows the laws of nature could in principle predict all future events (or facts) about the world. That is, determinism implies the possibility of perfect prediction. This possibility functions as an assumption in the Manipulation Argument and has been adopted unquestioningly *both* by incompatibilist defenders of the argument *and* by its compatibilist critics.[5]

The assumption should be rejected. Determinism does *not* entail the possibility of perfect prediction. As we will explain, because information cannot travel faster than the speed of light in our universe, determinism would *not* allow a predictor to predict with certainty any future event that is causally influenced by events outside her own backward light cone yet inside the backward light cone of the future event. As

[3] Recall, the *strong* incompatibilist thesis says that among the possible worlds whose laws of nature are deterministic, none contains free agents—or even stronger, in all possible worlds with (non-Humean) deterministic laws, those laws are the *reason* that no agents have free will (Mickelson, 2019). Campbell (2007) weakens this thesis by carving off those deterministic worlds that have eternal agents. Mele (2013) weakens it by carving off worlds whose deterministic laws are Humean. We maintain that if the actual world is deterministic, it does not permit deterministic manipulation. Hence, at best, the Manipulation Argument could establish only the *weak* thesis described above.

[4] Ismael (2013) argues that we ought to prefer a causal view of determinism over an entailment view.

[5] The claim that determinism entails perfect prediction is also an implicit assumption of other incompatibilist arguments and intuition pumps. To the extent that our argument undermines the assumption, it also weakens those arguments and the reliability of such intuition pumps.

a result, the Manipulation Argument, which relies on the assumption that determinism entails perfect prediction (and the ability to effectively manipulate on the *basis* of such prediction), cannot support the conclusion that determinism rules out human free will. At most, it supports the *weak* incompatibilist thesis that determinism might undermine free will in possible worlds with different laws of nature than ours, which would allow perfect prediction.[6]

Consider Danny, a normal agent in a deterministic universe who decides, at time t_1, to steal a wallet containing $100, which he finds on an empty street. Danny possesses all the capacities that compatibilists usually maintain would make him capable of acting freely and thus being morally responsible for his actions. Danny has the capacities to reflect on and identify with his desires (Frankfurt, 1971), to recognize and respond to reasons (Fischer & Ravizza, 1998), including moral reasons (Wolf, 1987), and he is not acting on compulsive or compelled desires (Mele, 1995). Combining these features (or others offered by compatibilists), Danny's decision to steal results proximally from his "Compatibilist Agential Structure," or CAS (McKenna, 2008)—i.e., from features of his psychology that compatibilists typically judge as jointly (and minimally) sufficient for free will.

Next, consider Manny, another agent in a deterministic universe. Manny differs from Danny only in that his decision to steal is causally determined by manipulators. For instance, in Pereboom's version of the Manipulation Argument (case 2, adapted), we are asked to imagine that "a team of neuroscientists programmed him [i.e., Manny] at the beginning of his life [t_0] … with the intended consequence that in his current circumstances [at t_1] he is causally determined to [steal the $100]" (2014: 77). In Mele's (2013) version, we are asked to imagine that Manny is created by a powerful goddess, Diana, who predicts that if she combines atoms in a particular way to create or alter a zygote, *Z*, at t_0, it will develop into Manny such that he will decide, using his compatibilist capacities (his CAS), to steal the $100 thirty years later at t_1: "From her knowledge of the state of the universe just prior to her creating *Z* and the laws of nature of her deterministic universe, she deduces that a zygote with precisely *Z*'s constitution" will produce her intended result (Mele, 2013: 175). From such cases, the Manipulation Argument follows:

> (MA 1): Because of the way he was designed in his deterministic universe, Manny lacks free will and is not morally responsible for deciding to steal the money.
> (MA 2): Regarding free will and moral responsibility for the decision to steal the money, there is no difference, in principle, between Manny and Danny.
> Therefore,
> (MA 3): Danny lacks free will and is not morally responsible for deciding to steal the money (and hence, free will and moral responsibility are incompatible with determinism; cf. Mele, 2006: 189).

[6] We do not concede that the argument works even if determinism permitted perfect prediction (Deery & Nahmias 2017). But we leave aside that objection.

If, as we will argue, determinism does *not* allow the predictions (and interventions) at t_0 that are required to intentionally design Manny such that he will decide to steal at t_1 in the way intended, then MA 1 fails to inform us about human free will. Hence, the conclusion does not generalize to Danny's action if Danny is supposed to be an agent in the sort of deterministic universe we might actually live in. As advocates of the argument themselves admit, when we consider possibilities (e.g., Humean laws) that "undermine [the] thought experiment" (Mele, 2013: 182), then "no version of the … argument can get off the ground" (Rogers, 2012: 292). If instead we focus on possible worlds in which determinism *does* allow perfect prediction, at best the thought experiment supports a conclusion that applies to agents in *those* worlds. We cannot generalize premise MA 2 or the conclusion MA 3 to agents like us in a world with a light-cone limit, even assuming our world is deterministic. We will return to these considerations in Sect. 4.

In the next section, we defend our claim that perfect prediction is not possible in any universe with our laws of nature, even if that universe is deterministic. Hence, Pereboom's neuroscientists and Mele's goddess Diana cannot know how to design Manny in a way that will ensure that he will decide to steal at t_1. In what follows, we will focus mostly on Mele's version of the argument.

3 The mistaken assumption

Following defenders of the Manipulation Argument, we characterize a perfect predictor as follows. Assuming determinism, a perfect predictor knows all of the laws of nature and knows the relevant states (typically the entire state) of the universe at an instant and thus she can calculate and predict with probability 1 what will happen at all later times (or some specific time). As Jenann Ismael (2019) observes, it is commonly taken to be an implication of determinism that such prediction is possible. In considering determinism within a Newtonian physical framework, Pierre-Simon Laplace famously wrote that:

> An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. (Laplace 1814/1951: 4)

This image of Laplace's Demon is also the image of Mele's Diana.

Assume determinism, as both defenders of the Manipulation Argument and their compatibilist opponents do. To be a perfect predictor, Diana must calculate by time t_0 what will happen 30 years later at t_1—i.e., she must be able to predict with probability 1 what will happen at t_1. Furthermore, for Diana to be an effective deterministic manipulator of the sort envisaged by the Manipulation Argument, she must be able to consider different counterfactual states of her universe at t_0 and predict with probability 1 what *would* happen at t_1 (30 years later) given those states, according to the laws. Whether Diana designs the zygote from scratch or alters an existing zygote, it is only because she makes a particular change to the zygote at t_0 that Manny will steal the money at t_1; otherwise, he (or some counterparts of Manny) would not. In addition to her ability to make such counterfactual predictions, Diana must be able to intervene physically to *implement* the required alteration to Manny's zygote at t_0, so as to ensure that Manny steals at t_1, 30 years later. Like Pereboom's neuroscientists who physically alter the infant's brain, Diana physically alters Manny's zygote.[7]

Now recall premise MA 1 of the Manipulation Argument:

> (MA 1): Because of the way he was designed in his deterministic universe, Manny lacks free will and is not morally responsible for deciding to steal the money.

This premise relies on an unstated assumption that the authors presume follows from determinism—namely, that the manipulators could, in principle, intervene to design Manny in the way described. If they could not, the argument would provide no reason to think that Manny lacks free will and is not responsible for deciding to steal—or at least no reason other than the assumption that determinism is true. Clearly, judging that Manny lacks free will for *this* reason alone would beg the question. The point of the Manipulation Argument is to present Diana's *manipulation* of Manny in order to motivate the judgement that he lacks free will, or to present determinism as entailing the possibility of such manipulation, and then to maintain—as premise MA 2 claims—that there is no difference relevant to free will between Manny and Danny. If Diana (or Pereboom's neuroscientists) cannot intervene as an effective deterministic manipulator, as premise MA 1 assumes, there is no reason to think that Manny lacks free will and is not blameworthy for deciding to steal, or at least no reason that advances the debate. As we will now explain, Diana cannot be such a manipulator if we are considering determinism as it might apply to humans, since perfect predictors are in fact nomologically impossible.[8]

3.1 The light-cone limit

Suppose that determinism is true in our universe and the laws of physics are relativistic.[9] In a relativistic deterministic universe, it is impossible for an agent like Diana, even with complete knowledge of the state of the universe at one time and the laws of

[7] To be an effective deterministic manipulator, the event at t_1 must *not* occur unless the manipulator *does* intervene—otherwise, all we have is perfect prediction (although we will argue that even perfect prediction is impossible). Pereboom and Mele both argue that merely predicting in order to ensure an outcome *without* intervening (as in Frankfurt cases) does not threaten free will, or at least not in the way manipulation or creation supposedly does in their arguments.

[8] In Sect. 1, we explained why incompatibilism is weakened if determinism must be supplemented in order to threaten free will. In Sect. 4, we will further address readers who object that nomological possibility is not what is important to these incompatibilist arguments.

[9] This assumption is the most plausible one to make about determinism in our universe. For reasons why a perfect predictor is similarly impossible under deterministic *Newtonian* laws, see Ismael (2019: 483).

nature, to predict with certainty any future event that is causally influenced by events outside her own backward light cone yet inside the backward light cone of the future event (we explain the notion of a light-cone in detail below). Because of the universal speed-limit set by the speed of light, there is a light-cone limit on how information can be acquired in our universe, such that Diana cannot acquire, at t_0, sufficient information to make perfect predictions about what will happen at t_1.

To grasp this idea, imagine that you have created a 5000-domino chain, such that domino #5000 will topple in exactly 500 seconds, triggering a switch that lights up a "Happy New Year" sign. If you topple domino #1 now (at time t_0), domino #5000 should switch on the sign at exactly midnight (time t_1). You topple domino #1 at t_0. Everything goes as planned until the last domino hits the switch. Yet the sign does not light up.

Why? Because the switch for the "Happy New Year" sign was hit by radiation from a solar flare, which left the sun just as you knocked over domino #1.[10] When the radiation reached Earth 500 seconds later, it disabled some electrical mechanisms, including the switch for your sign. Despite your control over the "deterministic" set-up of the dominoes, you could not have known about this effect of the flare when you toppled domino #1. The energy from the flare could not influence your predictions any faster than it actually arrived, at the speed of light. Thus, you could not have ensured the outcome that you intended and predicted at t_0.

Even a powerful being who has complete knowledge of all events about which information can have reached her, and who knows all of the laws of physics, cannot get information faster than the speed of light—at least if she is part of our universe.[11] Hence, even if our universe is deterministic, such a being cannot predict with certainty any future event that is causally influenced by events outside her backward light cone but inside the backward light cone of the future event.

Let us explain the light-cone limit in more detail, since the details matter. A Minkowski diagram of spacetime is a two-dimensional graph that represents space as one dimension and time as the other (see Fig. 1, which we model after a Minkowski diagram). Consider an event, *E*, that occurs at time t_0 (see Fig. 2, which makes clear the *cone* structure). For example, let *E* represent the radiation emitted from the sun at t_0 in our example. At t_1, this radiation will have travelled outward from *E* at 299,792,458 m per second, on the assumption that space is a vacuum. This speed (designated conventionally as *c*) is the maximum speed at which anything, including information, can travel in our universe.

Taking a time-slice of space at t_1, the distance the light has travelled since t_0 is represented by a circle (*C* in Fig. 2) whose diameter is twice the distance that light can travel between t_0 and t_1. Similarly, the circle representing how far the light has travelled by t_2 has a wider diameter, and so on, creating an increasingly larger "light cone." This is the forward light cone of *E*, which represents all the events in spacetime that could be causally influenced by *E*, while future events outside this forward

[10] In 1989, a large solar flare shut down electrical service to 6 million people for nine hours in Québec, Canada.

[11] Later, we will consider the possibility of *non-natural* predictors who are not part of the natural universe.

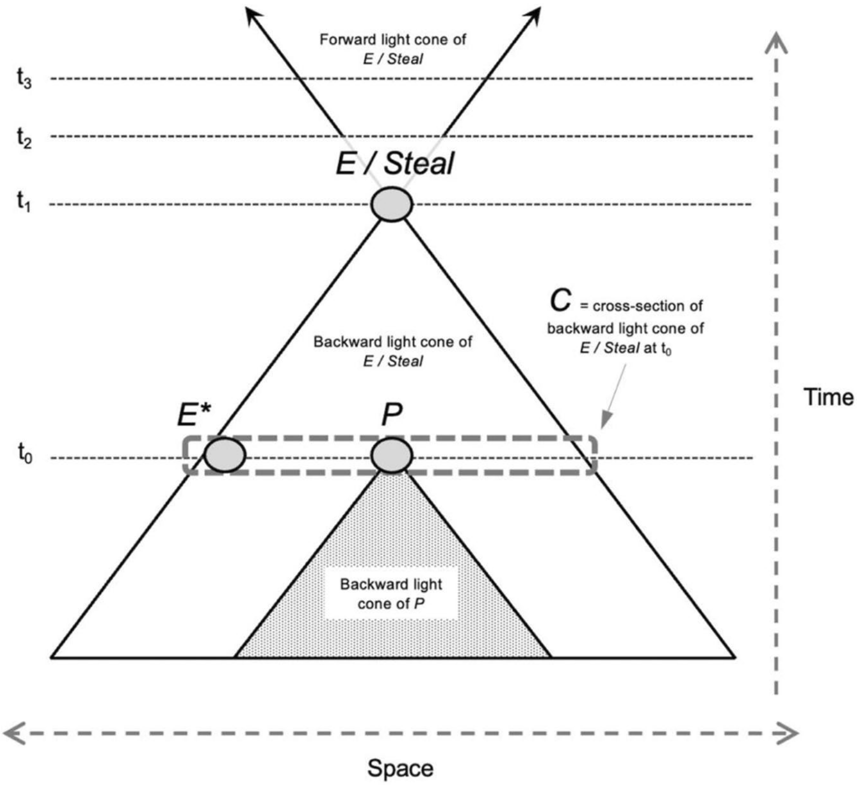

Fig. 1 Two-dimensional representation of light cone structure

light cone cannot be influenced by *E*. Additionally, *E* has a backward light cone, which represents all of the events prior to *E* that could causally influence *E*. Call a time-slice of *E*'s backward or forward light cone a "cross-section" of that cone (e.g., the cross-section *C* at t_1, in Fig. 2).

Standardly, there are three ways in which events on a cross-section of *E*'s backward light cone might be separated from *E* in spacetime. First, events on the surface of the cone—i.e., at the "edge" of any of its cross-sections—are *lightlike* separated from *E*, meaning that a photon from such events can causally influence *E*. Events inside (rather than on the surface of) *E*'s backward light cone are *timelike* separated from *E*, meaning that aphysical particle can travel from such events at a speed less than *c* and causally influence *E* (most causal processes are like this). What is important is the third way in which events in spacetime might be separated. To say that an event is *spacelike* separated from *E* means that no signal or particle can travel between that event and *E*, since doing so would require travelling faster than *c*.

Now imagine a powerful being attempting to make a prediction at t_0 about a future event *E* at t_1, where her prediction, *P*, occurs within *E*'s backward light cone (see Fig. 1). Even in a deterministic universe, the only way in which such a predictor

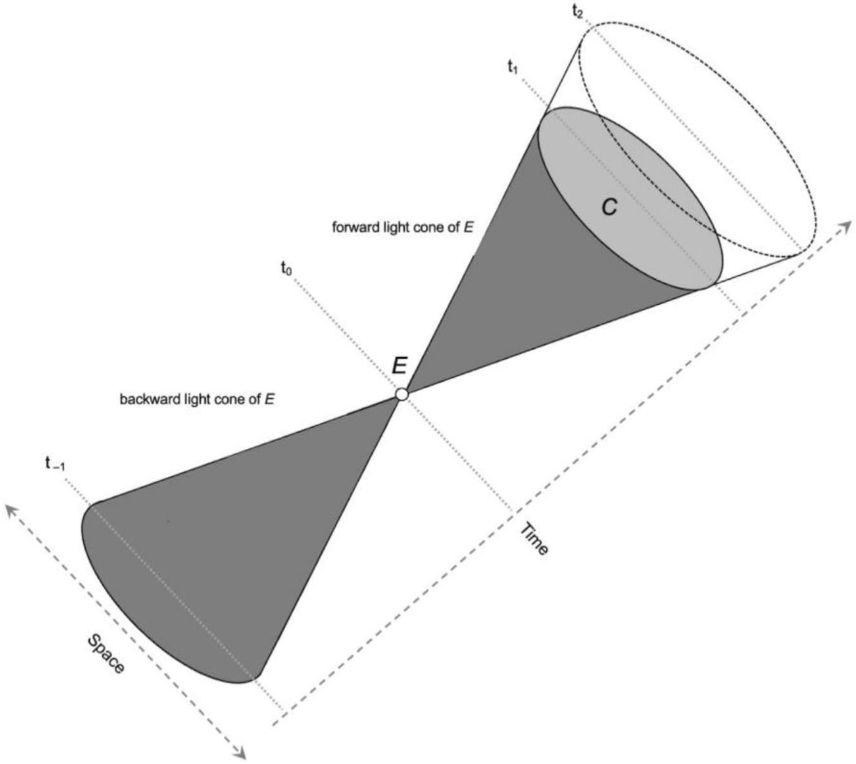

Fig. 2 Three-dimensional representation of light cone structure

could predict with probability 1 whether *E* will occur at t_1 (and thereby be a perfect predictor) would be if she knew all the laws of nature *and* had complete information about the events occurring within the cross-section, *C*, of *E*'s backward light cone at t_0. Otherwise, her prediction might fail if some event, *E** (Fig. 1), near the edge of *C*, initiated a causal process at t_0 that, propagating at the speed of light, impacted *E* prior to t_1. In our example of the dominos, if we take *E* to represent the switch for the "Happy New Year" sign being triggered by the toppling of domino #5000 at t_1, and *E** to represent the solar flare leaving the sun at t_0, *E** will influence *E* such that the flare's radiation disables the switch for the sign just prior to t_1, with the result that *E* does not occur—i.e., the sign does not light up. The predictor, predicting at *P*, while not timelike or lightlike separated from *E*, is nonetheless spacelike separated from events (such as *E**) that she would need to know about to perfectly predict whether *E* will occur.

The impact of the radiation on the switch cannot be predicted at t_0, even by a predictor who has perfect knowledge (at t_0) of *all* of the events from which she is not spacelike separated (at t_0). Hence, as perfect a predictor as there could be in our universe cannot perfectly predict all future events. Put another way, the information that a perfect predictor would need to know about at t_0 is, strictly speaking, information from her own future. From this being's perspective at t_0, there will be some future events she must know about to determine that event *E* occurs at t_1 that are not part of this being's causal past at t_0—indeed, some of these events will not be part of her causal past *until* t_1. As Ismael puts it, "As we get more future, we also get more past" (2019: 486). At t_0, the information about the future events that will influence *E* is simply unavailable; it will only be available to this aspirational oracle at t_1 (and thereafter).

This point is *not* merely epistemic. It is not that the relevant information about the events in the backward light cone of *E* is *already there* when the predictor makes her prediction at t_0 but she simply lacks access to this information. As Ismael puts it, "To have that thought … is to impose a conception of *the past* on Minkowski space-time that it doesn't support" (2019: 485). At t_0, numerous events in the backward light cone of *E* are located in what physicists call *the absolute elsewhere* for our predictor, since they are spacelike separated from her. Thus, the point is metaphysical: the relevant events do not exist for the predictor at t_0. As Ismael explains,

> The only meaningful spatiotemporal order in a Minkowski space-time is the causal order embodied in the light-cone structure. There's a well-defined order for events that fall in one's past causal horizon, and a well-defined order for events that fall in one's future causal horizon, but no well-defined order (relative to here-now) for events that fall in the absolute elsewhere. So there's no objective sense to be made of events that have *happened already*, but about which information isn't *here yet*. (2019: 485)

Despite the impact the Laplacean image of determinism has had on the way philosophers have understood it, determinism does *not* entail that all future events can be predicted based on the information available in the present. And to the extent that incompatibilist arguments, like the Manipulation Argument, depend on this image of determinism, they are misleading. As we now explain, a predictor such as Diana cannot make *perfect* predictions at t_0 (nor at any time significantly before t_1) about anything that will happen at t_1, including Manny's decision to steal.[12]

3.2 Diana and the light-cone limit

Let us represent with the variable *P* (for Predictor) the event of Diana's predicting (at t_0) what sort of zygote is required, such that it will develop into a person (i.e., Manny) who does what Diana wants 30 year later at t_1.[13] Let us represent the event of Manny's deciding to steal (via his CAS) the $100 at t_1 with the variable *Steal*, which

[12] The only exception would be future events that Diana could *completely* causally shield from any influence from events outside her own backward light cone at t_0. As even the highly controlled system in the dominos example suggests, complete shielding is impossible for events involving human agents over any extended period of time. See also footnote 14.

[13] The same points apply to Pereboom's neuroscientists predicting (at t_0) what sort of neural intervention is required at that time, such that it will lead to Manny's decision (from his CAS) to steal at t_1. Of course, the stipulation that human scientists, rather than a goddess, could gather all the information to ensure this outcome is implausible for reasons beyond those we present here.

is analogous to *E* in Fig. 1. For any would-be manipulator to ensure (at t_0) that Manny will steal, she must have perfectly accurate information about all the events in the cross-section *C*, at t_0, of the backward light cone of *Steal*. Otherwise, she cannot be sure what will happen at t_1, since it will be possible for an event like *E** at or near the edge of *C* to impinge on *Steal* in a way that would influence whether or how it occurs.

For instance, suppose that, outside of Diana's backward light cone—i.e., her past—at t_0, an alien civilization 30 light-years away from Earth sends a message containing information about its existence (event *E**). From Diana's perspective at t_0, this event is in the "absolute elsewhere" (Ismael, 2019: 484). There is no objective sense in which it exists for her.

Even so, were the message received on Earth shortly before t_1, it would influence local events in ways that might affect whether Manny steals; for instance, Manny might stay home watching the news about the aliens rather than be on the street finding the wallet. A predictor with no access to *E** (since *E** is spacelike separated from her) cannot predict at t_0 the effects *E** might have at t_1. Thus, she cannot take *E** (or its potential effects) into account when predicting Manny's decision at t_1, nor counterfactually predict with probability 1 how causal interventions at t_0 might *alter* Manny's decision. As a result, she cannot intervene at t_0 to ensure that *Steal* occurs in the way she intends. Again, this is a metaphysical not merely an epistemic point: the relevant events do not exist yet for Diana. The required information is too far away to have been received by Diana, even at the speed of light, at the time of her prediction, yet not too far away to influence the future event. Given the number of events in *C* that are spacelike separated from *P*, and about which the predictor cannot know, and given that events in complex systems (such as human brains and societies) can vary dramatically due to small differences in initial conditions, as suggested by non-linear dynamics, this is not an isolated example. The likelihood of failures in prediction increases dramatically with increased time between a prediction and the predicted event, even if all events are deterministically caused by prior events.[14]

As a result, no would-be manipulator, including Diana, can be a perfect predictor in a universe such as ours with a light-cone limit. Consequently, Diana cannot be an effective deterministic *manipulator* in any such universe.[15] As such, the Manipulation Argument fails as an argument for strong incompatibilism. It does not show that determinism rules out free will in universes like ours in which effective deterministic manipulation is impossible, because perfect prediction is impossible. If manipulation by a perfect predictor is supposed to illuminate the potential threat posed to human free will by determinism, it fails, since determinism, even if true in our universe, would not allow for such prediction.

[14] The light-cone limit minimizes possibilities for perfect prediction even on smaller time scales. Admittedly, Earth is a *relatively* closed system, so Diana *might* not fail for some near-future predictions. Similarly, mere humans can predict with relative confidence *some* near-future events, especially ones they control in real time by adjusting their actions to achieve the intended goals despite potential obstacles.

[15] For example, Diana (or Pereboom's neuroscientists) must know *how* to intervene to make changes that would result in the later events' occurring and must have the power to *implement* this knowledge to bring about those changes.

4 Responses to objections

The light-cone limit undermines the ability of the Manipulation Argument to establish the strong incompatibilist thesis. More importantly, the argument is weakened such that the condition it must add that *would* allow for perfect prediction does *not* apply in our universe. As a result, the argument cannot help to illuminate whether determinism in our world would rule out *human* free will. Below, we consider several objections to this claim.

4.1 Lucky diana

First, advocates of the Manipulation Argument might try to develop cases in support of premise MA 1 that do not rely on perfect prediction. For instance, they might present Diana as doing something at t_0 that merely happens to be one among the (enormous) set of causal conditions that deterministically result in Manny's stealing at t_1. For example, Diana might create a zygote she *hopes* will develop into a person, Manny, who will steal $100 in 30 years, and she might get lucky insofar as her doing so ends up being part of what actually causes Manny to steal $100 at t_1. Or perhaps Diana gets lucky because, as it turns out, no events she could not know about at t_0 actually end up influencing Manny during his entire lifetime before t_1.[16]

Relatedly, Pereboom (2014: 82) asserts that the intuition that Manny is unfree remains even if we replace intentional manipulators with a random physical event, such as a force field, which just happens to bring about the relevant changes in the agent. However, this move significantly weakens the Manipulation Argument by weakening the intuitive plausibility of premise MA 1. Here, the so-called "manipulation" is not really manipulation at all, since it is no different from regular causal influences in a deterministic universe. Whether determinism is true or false, various physical events like the ordinary development of Danny's zygote at t_0, represented by *Z*, will causally contribute to, and be necessary causal conditions for, Danny's decision 30 years later (e.g., *Z*'s occurring as it does might contribute to Danny's existing rather than some other person's existing). But *Z*'s occurrence does not causally control or *ensure* what Danny does in the way that a manipulator's intentional activity—represented by *M*—is presented as doing in the cases. *M*'s occurrence, unlike *Z*'s, is presented as ensuring that a particular decision occurs *rather* than some other decision, and as doing so across a wide range of possible background conditions—e.g., events Diana foresees that might causally interact with the decision and interfere with her plan. People are far less likely to have the intuition required to support premise MA 1 if the manipulator, like an ordinary physical event, cannot control

[16] A feature of these cases that is essential for the argument to work, but is typically left unanalyzed by its advocates, is that Manny must satisfy compatibilist conditions. So, the manipulator must not only create him so that he carries out her intended action 30 years later; she must do so in such a way that he acts through his CAS. Doing so would require that Diana cannot merely get lucky in contributing to the outcome but must tailor her design to ensure the action is caused in the right way. If Diana cares so much about Manny's acting in ways that satisfy compatibilist conditions, people may well have intuitions about Manny's free will more akin to those elicited by Frankfurt cases than those elicited by real-world manipulation.

Manny's decision in these robust counterfactual and contrastive ways, as the original manipulation cases stipulate.[17]

Unless Diana has the predictive powers required for such control, it is unclear why, in these weaker characterizations of Diana's "manipulation," we should (or most people would) think that Manny at t_1 lacks free will and is not responsible for deciding to steal. Hence, Diana needs to be a *counterfactual* predictor, so that she can consider which alterations to Manny's zygote at t_0, among many she considers, will produce the specific outcome she wants, i.e., Manny's stealing (from his CAS) at t_1. Once Diana knows which alteration will produce the desired outcome, she uses her powers to intervene accordingly. In this way, Diana has *counterfactual control* over the outcome.

Does Diana need such counterfactual control? Advocates of the Manipulation Argument might claim it is enough for Diana to be a significant cause in the distant past, even if she is not an intentional manipulator (Björnsson & Pereboom, 2016). We disagree. If Diana is not a perfect predictor (and thus is not a counterfactual predictor), it is unclear how she can have a relevant intention about anyone 30 years in the future stealing a wallet at a specific time. Instead, Diana might think, "I bet if I left this zygote alone, events would unfold in a particular way, and if I tweaked it, they would unfold in a different way. So, I'll tweak it and see what happens!" Thirty years later, seeing Manny steal the wallet, Diana exclaims, "Wow, the person into whom that zygote developed stole a wallet!" If the manipulator cannot even form an effective intention about her desired outcome, that should not count as manipulation at all (cf. Fischer, 2016).

A related objection maintains that perfect prediction is too demanding, since some manipulation arguments present the manipulation as occurring in an *indeterministic* universe (e.g., Cyr 2020). Such examples actually strengthen our argument. In some of these arguments, indeterministic universes are stipulated as allowing perfect prediction for only two types of outcomes—ones that the manipulator intends to happen (e.g., Manny's stealing the wallet) and others in which the agent loses agential capacities. The light-cone limit undermines both these stipulations. Without access to information outside her backward light cone at the time of her initiating an indeterministic causal chain, a manipulator could not predict how likely either of these two stipulated options is or control the relevant events.

In other indeterministic cases, the manipulator is presented as remaining "online" to observe and ensure that the manipulated agent will carry out the manipulator's intended goals (much like a Frankfurt intervener). This stipulation might enable Diana to ensure that Manny steals the wallet, whether indeterminism is true or a light-cone limit exists or both. But it also introduces an element to the manipulation that makes premise MA 2 much less plausible, since it suggests a type of fatalism that determinism does not entail. Determinism does not ensure that certain outcomes occur *no matter what* else might occur or no matter what an agent might try to do. If manipulators in these cases must observe Danny's life unfold and be prepared to

[17] For empirical support of these claims about people's intuitive judgments about manipulation and when it lowers judgments of freedom and responsibility, see e.g., Murray & Lombrozo (2017), Phillips & Shaw (2014).

intervene if it ever deviates from their intended plan, then they have a type of counterfactual control over him that does not illuminate, but distorts, our understanding of determinism, since regular determinism does not work like that. This type of online manipulation is not the direct target of our argument here, since the light-cone limit does not rule it out—if anything, it helps to illuminate *why* such online control *would* be required for the envisioned manipulation to be possible in our universe.

4.2 What sort of possibility matters for manipulation arguments?

Instead of giving up on Diana's ability to predict and be an effective deterministic manipulator rather than merely a lucky causal influence, incompatibilists might insist that it is irrelevant to the Manipulation Argument whether the manipulator in premise MA 1 is *nomologically* possible. Why not grant Diana the knowledge she would otherwise lack by putting her outside the natural universe and giving her complete knowledge that way, or by making her immune to the light-cone limit in another manner? This objection asks why we should restrict ourselves to considering nomologically possible cases instead of considering cases that are in some other way metaphysically possible or merely conceptually coherent.

As we explained in Sect. 1, this move requires stipulating conditions in addition to determinism to make the manipulation case work. One such condition stipulates that Diana exists outside the natural universe, yet as a non-natural godly agent can somehow intervene causally in the natural world. That requires first that Diana can somehow have complete knowledge of all the events in the cross-section at t_0 of the backward light cone of the event she wants to control at t_1. It also requires that she can intervene in the natural universe in order to effect the required manipulation. This latter condition introduces the *problem of interaction* that Princess Elisabeth of Bohemia first raised against Descartes and which has posed a challenge for interactionist dualism ever since (e.g., Shapiro, 2007: 64).[18] In short, it is unclear how Diana, even if she possessed the knowledge we maintain she would lack in our world, could intervene causally in any *physical* world. At the least, the incompatibilist owes us an explanation not just of these nomologically impossible conditions but also why they should be understood as illuminating the alleged threat of determinism to free will rather than introducing a distinct potential threat.

Relatedly, advocates of the Manipulation Argument might stipulate that Diana is the creator of the entire universe and can somehow foresee how various Big Bang creation events will deterministically ensure specific outcomes that she desires—perhaps all of them or merely some of them, such as Manny's deciding (using his CAS) to steal the wallet at t_1, 13.8 billion years later. The idea, we take it, is that stipulating Diana as an *initial* creator of the universe might allow her to escape the light-cone limit, or be in some way part of the physical universe, or somehow solve the interaction problem just once at the beginning of space-time. However, such stipulations seem to shift the discussion to debates about theological determinism

[18] See also Kim's "pairing problem" (2011: 50–54) and Dennett's argument that interactionist dualism violates the conservation of energy (1991: 35). Dennett's argument would apply even if Diana is physical but outside the closed physical system of our universe.

or pre-ordination. If advocates of the Manipulation Argument suggest that causal determinism threatens free will for the *same* reasons that theological pre-ordination might, we are happy to have pushed them to that position. But if the Manipulation Argument instead requires *supplementing* determinism with these theological conditions in order to develop manipulation cases, it has not shown that determinism *per se* threatens free will. If such theological conditions are ones that are unlikely to hold in a universe like ours, then the argument is weakened further, such that it cannot show that determinism by itself would threaten *human* free will. At best, determinism might be thought to threaten free will only in universes with different laws (without a light-cone limit) or in universes that have an intentional creator.

Similarly, recall that Mele himself concedes that his version of the Manipulation Argument does not work in any universe where the laws of nature are Humean (2006: 194–5; cf. Beebee & Mele, 2002). Since Humean laws are merely summaries or generalizations about how events have occurred, they do not govern (or permit predictions of how) future events *must* occur. So, even Diana, with perfect knowledge of the state, *S*, of the universe at t_0, could not predict with certainty that *S will* result in Manny's stealing at t_1, given the laws of nature. It is only on a non-Humean conception of the laws that a being such as Diana could make this prediction. Our argument adds that even on a *non-Humean* conception of laws, consideration of the light-cone limit demonstrates that Diana's prediction (and manipulation) is impossible in any world with our natural laws, even if these laws are non-Humean. At best, the Manipulation Argument might establish the incompatibility of free will and determinism in worlds with *different* non-Humean laws than our own, i.e., worlds without a light-cone limit.

We have shown that the Manipulation Argument is weakened by any of these attempts to avoid the light-cone limit either by giving up on the claim that the manipulator must be able to predict her intended outcome or by salvaging her predictive abilities by ignoring the limits imposed by the actual laws of nature.

5 Methodological considerations

We end by reflecting on some methodological considerations raised by our arguments, each of which poses difficult questions for advocates of Manipulation Arguments. First, in addition to the specific reasons we have provided above, there are more general arguments for the methodological claim that, when doing metaphysics, we should restrict ourselves to nomological possibility.

For example, Dorothy Edgington (2004) argues that there are two separate families of possibility, metaphysical and epistemic. Metaphysical possibility, Edgington claims, *should* be constrained by the actual laws of nature and so by what is nomologically possible, whereas logical possibility (which is roughly what many philosophers seem to think metaphysical possibility consists in) is best understood in epistemic terms. In brief, Edgington thinks Humean attempts to understand the laws of nature as contingent regularities do not explain how laws differ, as they seem to, from merely accidental contingent regularities. Instead, Edgington suggests we should think of laws as being in an important sense *necessary*: "Nothing *can* travel faster than light. These plants *can't* be grown at freezing temperatures. These other plants, merely, never are, in the history of the universe, grown at freezing temperatures, although they could have been" (Edgington, 2004: 3). In everyday speech and science, Edgington argues, metaphysical possibility *is* (and should be) thought of as constrained by the laws governing our actual universe. This general sort of non-Humean view about metaphysical possibility, as constrained by nomological possibility, has been widely defended recently, from philosophy of science to metaphysics, including by Andrea Borghini and Neil Williams (2008), Barbara Vetter (2013), and John Heil (2015).

Why not imagine a world in which Diana's manipulation of Manny is presented as possible, perhaps because Diana's knowledge is not subject to a light-cone limit? Because that world is not nomologically possible, and on the view that metaphysical possibility is constrained by nomological possibility, it is therefore not metaphysically possible, even if we can imagine it. As Edgington puts it:

> I do not mean to be mean-spirited about what possibilities there are. We can let our imaginations rip and speak of all manner of weird and wonderful possibilities. They are … epistemically possible. That is, they can't be ruled out *a priori*. We also need a more constrained notion: the possibilities for this world, and for the things that are in it, the various really possible histories they could have. (2004: 21)

Accordingly, Manny's history as including Diana's manipulation is *not* a metaphysically possible history. Rather, it is simply a case of letting our imaginations rip—a "weird and wonderful" possibility. On this view about metaphysical possibility, we should take seriously the requirement that any history for Manny must include a light-cone limit. If advocates of the argument want to argue that we should "let our imaginations rip" and consider nomologically impossible manipulation to illuminate the consequences of determinism, they should, at a minimum, acknowledge this methodological presupposition and ideally provide arguments in support of it.

Advocates of the Manipulation Argument might object that they are indeed asking us to let our imaginations rip—using epistemic possibility to help illuminate the nomological possibility of determinism. They might point out that they never meant to suggest that the existence of a goddess like Diana was as likely as determinism to be true of our universe, or that they were offering *plausible* conjectures about a world like ours. Rather, advocates are relying on this thought experiment to highlight a consequence of determinism—i.e., that it entails that our actions are determined by earlier events ultimately beyond our control. For instance, Pereboom argues that his manipulation case is "a vehicle for making the supposition of causal determinism salient in a way that effectively brings it to bear on these intuitions, judgments, and related emotions relevant to freedom and responsibility" (2014: 88). As such, the claim might be that nomological possibility is irrelevant to the argument. Rather, the cases merely need to be epistemically possible, or conceptually coherent, such that we can understand them in a way that helps us to understand the consequences of determinism.

But if *that* is the goal, the cases should not incorporate features that distort rather than illuminate how determinism would work in our universe. Since our universe *does* have a light-cone limit, the cases *do* incorporate such a distorting feature. So, the cases do not helpfully illuminate determinism or its consequences.

At this point, incompatibilists might further argue that it is not as though the light-cone limit could possibly bear on our conceptual expertise (or on participants' intuitions in experimental philosophy surveys) regarding the concept of free will or the conditions of moral responsibility. They might argue instead that we are *extremely conceptually competent* with applying the concept of manipulation. We agree, and it is likely that we may even have evolved to develop conceptions of agency and responsibility whose contours were informed by our experience and observation of, and resistance to, manipulation by others. Indeed, our need to track attempts at manipulation by conspecifics is plausibly a significant contributor to the evolution of human intelligence (e.g., Byrne & Whiten 1988). Developmentally, too, each of us must learn to recognize cases of manipulation and avoid them. Advocates of the Manipulation Argument might point out that their argument is drawing on that sort of conceptual expertise to reveal something we might not easily see otherwise—namely, that if our universe is deterministic, it is *just like* this other phenomenon, manipulation, which we competently understand as a threat to free will and responsibility.

There are several responses to make here. First, as our arguments show, Diana or Pereboom's neuroscientists do *not* illuminate the types of manipulation we *do* have conceptual experience or expertise with, since they are not only implausible but impossible in a world like ours with a light-cone limit. Since there are no manipulators of this sort in our world, we never encountered any in our evolutionary or individual learning histories. As a result, if we have the intuition that they undermine free will in virtue of their being initiators of deterministic causal chains, there is strong reason to think we are misapplying our concepts by *overreach*: by deploying them in ways unsupported by how we acquired them. What we require is an argument as to *why* these sorts of impossible cases *should* count as exemplars of the sorts of manipulation we have actually experienced.

Second, to the extent that we have conceptual expertise with the concept of manipulation, it is more plausible that it derives from the type of *control* we experience manipulators as exerting. Such experiences are based on manipulation being *different* from general causal histories, whether deterministic or indeterministic, not as illuminating how all causal histories work. While it is controversial how best to analyze manipulation, or whether any analysis can subsume all cases of manipulation (e.g., Greenspan 2003), most focus on features that undermine or bypass compatibilist capacities, such as reasons-responsiveness, or that involve real-time control, such that the manipulator can intervene if her dupe does not behave as she wishes.

We take it that on any plausible naturalistic view of how we come to acquire—through evolution and learning—various mental categories and concepts, it is typically *not* by considering logically possible yet nomologically impossible scenarios. Most plausibly, we acquire concepts such as those of free action (Deery 2021a, 2021b), agent (Sims 2019), or mental agent (Nichols 2017), *because* there is a relevant set of features in agents that we reliably track, which regulates our acquisition and retention of these concepts and serves important predictive or explanatory purposes for us. Whatever the details, one thing is certain. We do not track nomologically impossible features in our environment for the simple reason that *there are no such* *features* in our environment. So, if we are asked to decide whether a concept applies in a nomologically impossible case, it is unclear why we should trust our intuitive response to this question, since the case is different from any in which the concept *has ever* applied *or will ever* apply. Our expertise is insufficient to cover these cases.

Michael McKenna (2014) makes a related observation in responding to Pereboom's manipulation cases:

> It is reasonable to suppose that our intuitions have evolved along with our ordinary practices. Here, what I mean by intuitions are judgments in response to concrete cases expressing how pertinent concepts apply to them. It is my contention that the further away from ordinary contexts the application of these concepts are, the less reliable we should take them to be. There's a natural explanation for this. Our training for these concepts involved applications to contexts structured by our natural surroundings and our entire form of life. The more we move away from these, the less mooring we theorists have to feel confident that ordinary users are indeed applying our concepts properly. (2014: 479–80)

McKenna's claim is that Pereboom's (or Mele's) case is so far removed from ordinary contexts that our responses to it are plausibly unreliable, regarding whether concepts like those of free will, manipulation, and so on, apply to the case. By contrast, McKenna maintains that consideration of other cases that are closer to ordinary contexts, where determination of an agent's action "is not by … design … but by the vagaries of life" (2008: 156), will prompt us instead to respond that such agents *do* act freely, notwithstanding their being determined by factors beyond their control. Because McKenna *agrees* with premise MA 2 of the Manipulation Argument, which says that there is no difference relevant to being able to act freely between Manny and Danny, McKenna can conclude that Danny acts freely too (i.e., compatibilism is true). This is the so-called *hard-line* compatibilist response to the Manipulation Argument, which grants that there is no relevant difference between Manny and Danny yet denies that Manny is unfree because of the way in which he is manipulated by Diana. McKenna's claim that intuitive responses to cases like that of Diana are unreliable is bolstered by our claim that the envisaged manipulation is nomologically impossible and thus fails to highlight any consequence of determinism that would apply to human agents in our universe.

Finally, our argument offers a way to weaken the Manipulation Argument by putting direct pressure on premise MA 2, which claims there is no relevant difference between Manny and Danny. This is the so-called *soft-line response* to Manipulation Arguments, which grants that Manny is unfree yet insists that he is relevantly different from Danny (see e.g., Deery & Nahmias 2017; Schlosser 2015; Barnes 2013; Demetriou 2010). One simple way to make this move is to stipulate that it is relevant to free will and moral responsibility whether an agent can make decisions that are not perfectly (and perfectly counterfactually) predictable in such a way that they could be subject to effective deterministic manipulation by another agent (or, at least, by another agent within our physical universe). If so, then an agent such as Danny who exists in a universe like ours with a light-cone limit, even if that world is indeed deterministic, would differ in this way from Manny once we recognize that Manny is subject to such manipulation only because his universe has no light-cone limit. Consequently, premise MA 2 of the Manipulation Argument would be false. More generally, the light-cone limit helps us to see that even if our universe is deterministic, our future decisions are not even in principle perfectly predictable.

6 Conclusion

According to our best theories of physics, prediction has a light-cone limit, such that it is not possible to predict with certainty an event that will happen at a specific time far in the future. As a result, no manipulator in a universe like ours—even if it is deterministic—can know enough at t_0 to intentionally ensure that an alternative event occurs at t_1 instead, as the Manipulation Argument stipulates in support of the claim that Manny is unfree—premise MA 1—before asserting that Danny is no more free or responsible than Manny—premise MA 2. Our argument forces advocates of the Manipulation Argument to give up the strong incompatibilist thesis that, necessarily, free will is impossible if determinism is true. At best, the argument might support a weak incompatibilist thesis that free will is impossible if both determinism *and* a condition that does not apply in our universe were true. Our argument also points to potentially relevant differences between Manny and agents like Danny in universes like ours.

Additionally, we have provided several methodological considerations that, at a minimum, put the ball back in the court of advocates of the Manipulation Argument, requiring them to defend their use of this nomologically impossible thought experiment, their understanding of conceptual competence, and their understanding of strong versus weaker forms of incompatibilism. We may lose some of our audience of philosophers (incompatibilist or compatibilist) who *accept* the methodological suppositions that might permit use of the stipulations required for premise MA 1 to work. Nevertheless, as those engaged in these debates have recognized, the primary audience is not defenders of the Manipulation Argument or even compatibilists who engage with the argument on their own terms. Instead, the target audience is those who have not yet made up their mind on the question of whether free will is compatible with determinism. We have offered members of that audience reasons to be wary of the background assumptions required to prop up the Manipulation Argument.

More generally, we have shown that the possible truth of determinism does not entail that human decisions could be perfectly predicted, even in principle, nor that a manipulator could ensure what we do. To the extent that any incompatibilist arguments or intuitions are influenced by this understanding of determinism, they mislead us. Determinism might be true in our universe. But if it is, its implications are very different from what many have imagined them to be. By the same token, we might lack free will. But if so, it is likely because of more pressing threats than the potential truth of determinism (e.g., Nahmias 2007, 2014; Levy 2016; Deery 2021a; Bernstein forthcoming). Even if we lack free will *because* determinism is true in our world, it is *not* because determinism implies the possibility of perfect prediction or manipulation. If we must consider determinism in relation to free will, let us at least get the implications of the thesis right.

Acknowledgements This paper began over 7 years ago (and, as a result, early versions of the manuscript emailed out at the speed of light may now be influencing intelligent life on planets located 7 light years from Earth). In the process, we have received useful comments from far too many people to remember. But we are pretty sure that some of them include Joe Campbell, Jenann Ismael, Jason Turner, Terry Horgan, Scott Collison, and Al Mele. We would also like to thank some but not all of the referees of previous versions of the paper and we are deeply grateful to the referees of this version. We do apologize for failing to remember any others who helped us on this multi-year odyssey.

References

Barnes, E. C. (2013). Freedom, Creativity, and Manipulation. *Noûs*, *49*(3), 560–588

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

Bernstein, S. (forthcoming), & Shoemaker, D. (Eds.). “Resisting Social Categories,” in *Oxford Studies in Agency and Responsibility, Vol. 6*, D. Shoemaker, Oxford:Oxford University Press

Björnsson, G., & Pereboom, D. (2016). “Traditional and Experimental Approaches to Free Will and Moral Responsibility. In J. Sytsma, & W. Buckwalter (Eds.), ” *Companion to Experimental Philosophy* (pp. 142–167). Oxford: Blackwell Press

Borghini, A., & Williams, N. (2008). A Dispositional Theory of Possibility. *Dialectica*, *62*(1), 21–41

Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Oxford University Press

Campbell, J. (2007). Free Will and the Necessity of the Past. *Analysis*, *67*, 105–111

Campbell, J. (2008). Reply to Brueckner. *Analysis*, *68*, 264–269

Deery, O., & E. Nahmias. (2017). “Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood,” *Philosophical Studies*, 174(5): 1255–1276.

Deery, O. (2021a). *Naturally Free Action*. Oxford: Oxford University Press.

Deery, O. (2021b). “Free Actions As a Natural Kind,” *Synthese*, 198: 823–843

Demetriou, K. (2010). The Soft-line Solution to Pereboom’s Four-case Argument. *Australasian Journal of Philosophy*, *88*(4), 595–617

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.

Edgington, D. (2004). “Two Kinds of Possibility,” *Proceedings of the Aristotelian Society*, 78(1): 1–22

Fischer, J. (2016). “How Do Manipulation Arguments Work? *The Journal of Ethics*, *20*(1–3), 47–67

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, *68*, 5–20

Greenspan, P. (2003). The Problem with Manipulation. *American Philosophical Quarterly*, *40*(2), 155–164

Heil, J. (2015). “Aristotelian Supervenience,” *Proceedings of the Aristotelian Society*, 115(1): 41–56

Ismael, J. (2013). “Causation, Free Will, and Naturalism. In H. Kincaid, J. Ladyman, & D. Ross (Eds.), ” *Scientific Metaphysics* (pp. 208–235). Oxford: Oxford University Press

Ismael, J. (2019). Determinism, Counterpredictive Devices, and the Impossibility of Laplacean Intelligences. *The Monist*, *102*, 478–498

Kim, J. (2011). *Philosophy of Mind*. New York: Taylor & Francis

Laplace, P. S. (1814/1951). *A Philosophical Essay on Probabilities*, trans. F. W. Truscott and F. L. Emory. New York: Dover Publications

Levy, N. (2016). Implicit Bias and Moral Responsibility: Probing the Data. *Philosophy & Phenomenological Research*, *94*, 3–26

McKenna, M. (2008). A Hard-line Reply to Pereboom’s Four-case Manipulation Argument. *Philosophy & Phenomenological Research*, *77*, 142–159

McKenna, M. (2014). Resisting the Manipulation Argument: A Hard-liner Takes It on the Chin. *Philosophy & Phenomenological Research*, *89*, 467–484

Mele, A. (1995). *Autonomous Agents: From Self-control to Autonomy*. New York: Oxford University Press

Mele, A. (2006). *Free Will and Luck*. New York: Oxford University Press

Mele, A. (2013). Manipulation, Moral Responsibility, and Bullet Biting. *Journal of Ethics*, *17*, 167–168

Mele, A. (2019). *Manipulated Agents: A Window to Moral Responsibility*. New York: Oxford University Press

Mickelson, K. (2019). The Problem of Free Will and Determinism: An Abductive Approach. *Social Philosophy and Policy*, *36*, 154–172

Murray, D., & Lombrozo, T. (2017). Effects of Manipulation on Attributions of Causation, Free Will, and Moral Responsibility. *Cognitive Science*, *41*(2), 447–481

Nahmias, E. (2007). “Autonomous Agency and Social Psychology,” in *Cartographies of the Mind: Philosophy and Psychology in Intersection*, ed. Marraffa, Caro, and Ferretti (pp. 169-185). New York: Springer Press.

Nahmias, E. (2014). “Is Free Will an Illusion? Confronting Challenges from the Modern Mind Sciences,” in *Moral Psychology, vol. 4, Free Will and Moral Responsibility*, ed. W. Sinnott-Armstrong (pp. 1-25). New York: MIT Press.

Pereboom, D. (2001). *Living Without Free Will*. New York: Cambridge University Press

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press

Phillips, J., & Shaw, A. (2014). Manipulating Morality: Third-Party Intentions Alter Moral Judgments by Changing Causal Reasoning. *Cognitive Science*, *38*(8), 1320–1347

Rogers, K. A. (2012). The Divine Controller Argument for Incompatibilism. *Faith and Philosophy*, *29*, 275–294

Shapiro, L. (Ed.). (2007). *The Correspondence Between Princess Elisabeth of Bohemia and René Descartes*. Chicago: The University of Chicago Press

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Oxford University Press

Vetter, B. (2013). ‘Can’ Without Possible Worlds: Semantics for Anti-Humeans. *Philosophers' Imprint*, *13*(16), 1–27

Warfield, T. (2000). Causal Determinism and Human Freedom are Incompatible: A New Argument for Incompatibilism. *Noûs*, *34*, 167–180

Wolf, S. (1987). “Sanity and the Metaphysics of Responsibility. In F. Schoeman (Ed.), ” *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 46–62). Cambridge: Cambridge University Press

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.