Episteme (2025), 1–16
doi:10.1017/epi.2025.8

episteme

ARTICLE

Simulation & Manipulation: What Skepticism (Or Its Modern Variation) Teaches Us About Free Will

Z. Huey Wen

New York University, New York, USA
Email: huey.z.wen@gmail.com

(Received 5 September 2024; revised 21 November 2024; accepted 3 January 2025)

Abstract

The chemistry of combining the simulation hypothesis (which many believe to be a modern variation of skepticism) and manipulation arguments will be explored for the first time in this paper. I argue: If we take the possibility that we are now in a simulation seriously enough, then contrary to a common intuition, manipulation very likely *does not* undermine moral responsibility. To this goal, I first defend the structural isomorphism between simulation and manipulation: Provided such isomorphism, either both of them are compatible with moral responsibility, or none of them is. Later, I propose two kinds of reasons – i.e., the simulator-centric reason and the simulatee-centric reason – for why we have (genuine) moral responsibilities even if we are in a simulation. I close by addressing the significance of this paper in accounting for the relevance of artificial intelligence and its philosophy, in helping resolve a long-locked debate over free will, and in offering one reminder for moral responsibility specialists.

Keywords: Simulation hypothesis; skepticism; virtual reality; manipulation argument; hard-line reply; moral responsibility

1. Introduction

"We are and always have been," says the simulation hypothesis, "in an artificially designed computer simulation of a world" (Chalmers 2022: 29).[1] Many philosophers treat this hypothesis not merely as a skeptical possibility but as a picture of the reality we are in.[2] Optimistically, if we are now in a simulation/virtual reality (two terms to be used interchangeably), many valuable insights concerning various questions in traditional philosophy will follow (Chalmers 2022: 17–8). Nevertheless, virtual reality

[1]Broadly construed, the simulation hypothesis can also be understood as a modern variation of skepticism (about external world). In this vein, my later conclusion ("If the simulation hypothesis is true, then *p*") may in principle generalize to something like "If we are brains in vat, then *p*," or "If skepticism about external world is true, then *p*." Yet to entertain the timely discussion of AI and VR, I primarily focus on the simulation hypothesis.

[2]See, e.g., Bostrom (2003, 2009), Lewis (2013), White (2016), and Chalmers (2022). For the sake of argument, I assume the truth of this hypothesis in what follows. For a discussion of the opposite view, see, e.g., Harris (2024).

seems to have little to say about free will. "Free will is much less of a problem in an ordinary virtual world," writes Chalmers, "[i]f we have free will in ordinary physical reality, then we can equally have free will in virtual reality" (2022: 320).[3] I think Chalmers is roughly right in the sense that the intersection of simulation and free will does not yield any shocking conclusion like "If the simulation hypothesis is true, then libertarianism is false." However, I believe the simulation hypothesis still has a provocative lesson to teach us about free will and moral responsibility.[4] The lesson is, more specifically, about manipulation arguments, central to which is "the intuition that, due to the nature of the manipulation, the agent does not act freely and is not morally responsible for what she does" (McKenna and Pereboom 2014: 162).[5] In the last two decades, the most prominent debate around manipulation arguments is arguably between *hard-liner hard determinists* and *hardliner compatibilists*, both holding that there is no *relevant difference with respect to whether the agents in given cases are morally responsible for their actions* ('MR-relevant difference' henceforth for short) between manipulation and (causal) determinism. Provided the intuition that manipulation undermines moral responsibility, *hard-liner hard determinists* argue, *contra* the compatibilist idea that there could be moral responsibility (and free will) in a (causally) deterministic world, determinism undermines moral responsibility, too (and compatibilism is therefore false).[6] In defense of compatibilism, *hard-liner compatibilists* argue that since there is no MR-relevant difference between determinism and manipulation, and we intuitively believe agents in a deterministic world are morally responsible, we also seem to have good reasons to believe agents under manipulation are morally responsible, too (and compatibilism might still be true).[7] Dialectically, we have the former's *modus ponens* vs. the latter's *modus tollens*. Since both sides start from their intuitions (that clash), there seems to be little space for further argumentation: By far, neither side can successfully discredit the intuition the other side appeals to.[8]

[3]This view can be further strengthened to one concern which, as one anonymous reviewer pressed to me, goes like this: "The kind of isomorphism/similarity between simulation and manipulation seems almost too obvious. The simulation hypothesis can simply be understood as a picture of what 'our' world is (metaphysically) in the first place. So, if compatibilism is true of reality, it holds irrespective of whether we are simulated—and so why does this paper need to be written?" Presumably we are on the sample page regarding the "no difference" conclusion, and the reviewer walks on a quick path toward it. Their path, however, is arguably not the most rewarded (nor the most insightfully) one. It misses the opportunity of cutting the concept of manipulation (and simulation) open, figuring out what these key elements we care about when we talk about them are, and most importantly, insights on their nature. In other words, my path, by contrast, signifies not (only) because of its conclusion but, more importantly, because of its process.

[4]Notably, free will is usually believed to be a necessary condition (aka the control condition) of moral responsibility, see, e.g., Mele (2006: 27 n. 18), van Inwagen (2008: 329), and McKenna and Pereboom (2014: 6).

[5]See, e.g., Kane (1996: 65), Pereboom (2001, 2014), Liu (2022), and De Marco (2023).

[6]Derk Pereboom (2001, 2014) is of course the most famous defender of this position. Another advocate of hard-liner hard determinism is arguably Kristin Mickelson (2015), who defends an even stronger position: There is no MR-relevant difference between manipulation and determinism, therefore free will and moral responsibility is impossible for agents living in a deterministic world (like us).

[7]Micheal McKenna (2008, 2014) is of course the most famous defender of this position. Other advocates of hard-liner compatibilism include, e.g., John Martin Fischer (2011), and Sofia Jeppsson (2020).

[8]Notably, there is another line of response in the literature (aka *the soft-line reply*) that aims at articulating what the MR-relevant difference between determinism and manipulation is. Philosophers have developed two such accounts: The bypassing view and the manipulator-focused view (De Marco 2023: 487). The bypassing view suggests the relevant difference concerns the fact that the action issues from attitudes that the manipulated agent acquired in a way that bypassed her capacities for control over her mental life, whereas the manipulator-focused view suggests the difference concerns the presence of a manipulator. For soft-liner replies in the recent literature, see, e.g., Barnes (2013), Deery and Nahmias (2017), and Usher (2018).

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

However, with resources from the discussion of simulation, I might have an argument showing how the intuition (which hard-liner hard determinists appeal to) that "Manipulation undermines moral responsibility" is probably false.[9] My master argument runs as follows[10]:

1. If we are in a simulation, then we are manipulated.
2. Even if we are in a simulation, we still have moral responsibilities.
3. We are in a simulation. (*The simulation hypothesis*)
4. Therefore, manipulation is compatible with moral responsibilities. (1, 2, 3)

This paper shall further proceed like this. § 2 justifies premise 1, in which I argue there is no MR-relevant difference between simulation and manipulation. That said, simulation and manipulation are isomorphic by virtue of manifesting the same structure, and alleged differences between simulation and manipulation, if any, are arguably moral responsibility-irrelevant. § 3 justifies premise 2, of which I will provide two kinds of reasons in support. The simulator-centric reason says simulators should allow inhabitants of simulations (also referred to as "simulatees" or "sims") to have moral responsibilities for the sake of meaningful lives. The simulatee-centric reason says simulations satisfy key external conditions of moral responsibility-susceptibility (henceforth "MR-susceptibility" for short) and therefore probably do not undermine moral responsibilities. § 4 first responds to the objection that my position arguably entails a kind of skepticism about moral responsibility – that is, since everybody in a simulation is manipulated, probably nobody should ever be held morally responsible for any action at all – which is dubiously question-begging. After that, I conclude by marking the three-fold significance of this paper in accounting for the relevance of the philosophy of artificial intelligence, in helping resolve a long-locked debate over free will, and in offering one reminder for moral responsibility specialists.

2. Another "no relevant difference" claim

The central thesis of § 2 says: There is no MR-relevant difference between simulation and manipulation. In defense of this thesis, I first argue that simulation and manipulation are structurally isomorphic (§ 2.1), and then demonstrate why some alleged differences between simulation and manipulation are moral responsibility-irrelevant (§ 2.2).

2.1. Highlighting the structural isomorphism

Before I dive into details, I mark one important rationale for introducing and defending the proposed isomorphism. Exegetically, there seems to be an initial tendency to move from manipulation in ordinary cases to manipulation in simulation. In the original Frankfurt-style Case, the manipulator manipulates one particular action of one particular agent, and Frankfurt (1969) asks: What our judgment about the agent's moral responsibility (or the absence thereof) should be? After that, Mele and Robb (1998) further ask: Is there any change regarding our judgment about the agent's moral responsibility (or the absence thereof) if the manipulator manipulates all actions of the

[9]If we replace "moral responsibility" with "autonomy" or "freedom" (cf. fn. 4 for the connection between these notions), then the statement will have even more advocates. Sophie Gilbert recently surveys through a vast body of literature and reports (2023: 360; footnotes omitted): "There is nothing more pervasive in the philosophical literature on manipulation than the claim that manipulation 'undermines,' 'subverts,' 'diminishes,' 'infringes,' 'threatens,' 'interferes with,' or 'disrespects' autonomy, at least much of the time."

[10]My thanks to David Chalmers for proposing a reconstruction of the argument in conditionals.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

particular agent? In a similar vein, I want to ask: Is there any change regarding our judgment about an agent's moral responsibility (or the absence thereof) if the manipulator manipulates all actions of all agents? To this extent, § 2.1 answers to an unexplored tendency in the relevant literature from Frankfurt-style Cases via Global Frankfurt-style Cases to simulations.

To begin, I'd like to demarcate what I mean by "simulation" and "manipulation," as there are various readings of either concept in the rich literature. The idea of simulation ultimately goes back to computer science, in which a simulation is usually defined as a program using step-by-step methods to explore the approximate behavior of a mathematical model that is run on a computer.[11] On this conception, the following features are arguably essential to a simulation *qua* a computer program. First, a simulation is pre-programed in the sense that it starts with some initial parameters set by the simulator.[12] Consider a (toy) simulation of a car accident for example. The simulator has to input some data like the vehicle's initial velocity, the friction of the road surface, etc., before the simulation is run. Second, after the initial parameters are set, the following simulating process nevertheless proceeds in a seemingly autonomous manner.[13] Here is how. Again, consider the simulation of a car accident. Suppose the crush in the simulation is about to take place at time *t*. Intuitively, the occurrence of the crush at time *t* is the result of the system's own evolvement given these initial parameters set prior to *t*. That said, provided these pre-programed parameters, the crush naturally follows without any need for the simulator's further intervention at *t*. To this extent, a simulation seems to manifest a kind of autonomy.

Bearing these two features of simulation in mind, I continue demarcating the concept of manipulation. There are obviously abundant possible conceptions and distinctions made about manipulation in the literature. However, what is directly relevant to our discussion is just a tiny tip of the iceberg. Quite remote from manipulations common to ordinary life in the form of guilt trips, gaslighting, peer pressure, negging, or emotional blackmail, the kind of manipulation discussed by free will/moral responsibility specialists is instead a highly artificial kind that is "imagined as happening via decidedly extra-ordinary methods" (Noggle 2022: § 1). My discussion in this paper only focuses on this artificial kind.[14] This kind of manipulation "typically includes a powerful agent [i.e., the manipulator] who is capable of altering her dupe's [i.e., the manipulatee] external or internal conditions to ensure that the dupe does what she wants" (Deery and Nahmias 2017: 1258). More particularly, two points about this kind of manipulation are worth marking. First, at least for hard-liners, the means of manipulation is arguably irrelevant to its manipulative nature. Manipulation could be done by directly interfering with the manipulatee's mental states by sending signals to affect her brain (e.g., Pereboom's *Case 1*), or via disposing the manipulatee to a designed deliberation outset such that she could only act in the way the manipulator wants (e.g., Pereboom's *Case 2*), or via the manipulator's pre-design in the (remote) past that causally necessitates the manipulatee's choice that aligns with the manipulator's purpose

[11]This "narrow" definition comes from *The Stanford Encyclopedia of Philosophy* entry of "Computer Simulations in Science," and it could be, though unrelated to our discussion, expanded for other needs (Winsberg 2022: §§ 1.1–3).

[12]My thanks to Xiaofei Liu for pushing me to articulate and defend this feature of simulation.

[13]My thanks to Peter Finocchiaro for challenging my initial conception of simulation by this reading of his.

[14]Though I agree the ordinary kind is very much worth pursuing and, as the author of *The Stanford Encyclopedia of Philosophy* entry of "The Ethics of Manipulation" laments, "[s]o far, few people have explored the connections between ordinary manipulation and the forms of global manipulation discussed in the free will literature" (Noggle 2022: § 1).

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

(e.g., Mele's Zygote case). Second, in manipulation, the manipulatee still seems to act through her own *Compatibilist Agential Structure*, i.e., features of her own psychology that compatibilists typically judge as jointly (and minimally) sufficient for free will and moral responsibility.[15] In other words, even if the manipulatee's decision inevitably aligns with the manipulator's purpose, from the manipulatee's own perspective, she still has control over her choice and action. For a clear presentation of this feature, recall Pereboom's *Case 2* and Mele's Zygote case: In either case, via her pre-designing, the manipulator does not directly bypass any of the manipulatee's relevant capacities at, or before, the time of action.[16]

By now, it is not hard to see the similarity between simulation and manipulation: They both manifest the same structure (call it *the key structure*). Consider the following contrast for illustration:

- *Simulation:* [Input A as pre-programming] **[Computing rules]→** [Output B as the result of the system's own evolvement]
- *Manipulation:* [Interference C as pre-design] **[Laws of causality]→** [Performance D as the result of the agent working through his *Compatibilist Agential Structure*]
- *The key structure:* [Prior state set by party E] **[Independent rules/laws]→** [Post state realized by party F that seems to bestow autonomy]

That said, both simulations and manipulations manifest the same structure composed of five key elements, as illustrated in the chart:

Chart. The Structural Isomorphism Between Simulation and Manipulation.

Elements of *the Key Structure*	Where/How the Element is Manifested in Simulation	Where/How the Element is Manifested in Manipulation
#1 Two parties	The simulator and the simulatee	The manipulator and the manipulatee
#2 A prior state set by one party	The simulator's pre-programming	The manipulator's pre-design
#3 Some independent rules/laws	Mathematical and logical theorems	Laws of nature (especially laws of causality)
#4 A post state realized by another party	The (approximation of) the occurrence of the simulated event	The manipulatee's performing the action the manipulator wants her to perform
#5 A special connection between the prior state and the post state via these independent rules/laws	Provided the prior state and these independent rules/states, the post state will be necessitated (i.e., will be realized with the probability 1)	

[15]The term "Compatibilist Agential Structure" comes from McKenna (2008: 142) and an observation similar to my point is also made by Deery and Nahmias (2017: 1258), though Deery and Nahmias themselves defend a soft-liner view that even if both a manipulated agent and an unmanipulated agent seem to work through their own *Compatibilist Agential Structures*, provided a particular view (i.e., interventionist) about causation, the causal sources of their actions can still be different. See also Tierney and Glick (2020) for a critique of Deery and Nahmias (2017).

[16]But some could say, like in Mele's Zygote case, the agent is still wronged in the sense of being treated as a mere means by the divine interferer (see Herdova 2021). However, notably, though this line of reasoning might be true, it is soft-liner in spirit.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

In short, by manifesting *the key structure*, simulation is isomorphic with manipulation.[17] By virtue of such structural isomorphism, I argue, there is no relevant difference between simulation and manipulation regarding whether, in their presence, an agent should be held morally responsible for her actions.[18] That said, any reason to hold an agent for (or exempt her from) moral responsibilities in cases of manipulation also applies to cases of simulation, and *vice versa*.[19] I only mark two examples of such reasons here. First, a hard-liner compatibilist might hold that agents under manipulation should be deemed morally responsible because any action she performs is still the result of her working through her *Compatibilist Agential Structure* and no relevant capacities of her are ever bypassed in the process. Similarly, a parallel reason could say an agent in simulation is also morally responsible in the sense that any action performed by her also seems to be the result of her working through a structure of her own psychology which is similar to *Compatibilist Agential Structure* in cases of manipulation. Second, on the contrary, a hard-liner hard determinist may hold that agents under manipulation should be exempted from moral responsibility because, despite the dubious autonomy they manifest, their later action is causally necessitated by the manipulator's pre-design, which arguably undermines moral responsibility. Similarly, a parallel reason could say an agent in simulation is also not morally responsible because her later action is necessitated by the simulator's pre-programming, too. That said, if manipulation undermines moral responsibilities, then so should simulation; and if simulation is compatible with moral responsibilities, then so should manipulation.[20]

I deal with one objection at the end of § 2.1. Consider the following depiction of the dialectical landscape:

[Place 1] — [Manipulation] — [Place 2] — [Determinism] — [Place 3]

One may say: Suppose simulation is compatible with moral responsibility; for your master argument to work, the case of simulation should presumably be added to Place 1; however, since simulation also resembles (causal) determinism, we may put it in Place 2 or 3, too; and if so, then hard-liner hard determinists may still reasonably hold the intuition that "Manipulation undermines moral responsibility," and the purpose of your paper is jeopardized.

[17]This isomorphism thesis should be distinguished from the thesis of simulation realism (as in, e.g., Chalmers 2022) that simulated realities are no different from/as real as/isomorphic with non-simulated realities, at least when it comes to these salient aspects of reality that we care. My isomorphism thesis is primarily a world-case relation that stands between simulated realities and manipulation cases (in non-simulated realities). The thesis of simulation realism is, by contrast, mostly a world-world relation. Though I acknowledge that the world-world relation might be connected to the world-case relation in some way, the discovery of such connection would better be left for further research.

[18]Though I agree there could be moral responsibility-irrelevant differences between simulation and manipulation despite such isomorphism. For example, given manipulator's particular manipulative intention (e.g., having Plum kill White), their credence of the presumed post state being true is arguably higher than a simulator's credence, since the latter is often uncertain of the approximated results (e.g., the influence of the melting of glaciers).

[19]Though there might be disagreement about whether there is any agent in a simulation at all. See my discussion in § 4, especially fn. 35. Let us now assume that simulatees are agents for the sake of my "no difference" argument.

[20]I will later demonstrate several reasons suggesting why there is probably moral responsibility in simulation in § 4. And notably, my reasons are proposed from perspectives neutral about the compatibilism/hard determinism debate at present, which arguably will not risk trivializing the debate.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

My reply is this. I agree that simulation, manipulation, and determinism are similar in the sense that, in all of them, the prior state and the rules/laws taken together ensure the inevitable realization of the post-state.[21] However, one apparent difference is that the prior state for simulation and manipulation is set up by simulators or manipulators; while the prior state for determinism is probably some events in the natural history (e.g., the occurrence of the Big Bang), or at least determinism is silent about where the prior state comes from.[22] Thus, because of the presence of a second party (i.e., a simulator or a manipulator), simulation is more similar to manipulation than to determinism and it probably should not be put in Place 3.

As for whether simulation belongs to Place 1 or 2, let's continue considering Pereboom's transitional *Cases 2 and 3*. Arguably, Pereboom's *Cases 1, 2, and 3* are all manipulative; though the closer it gets to *Case 4*, the more disguised and ambiguous the means of manipulation becomes. The manipulation in *Case 1* is done by directly controlling the manipulatee's mind. The manipulation in *Case 2* is done by preprograming the manipulatee's deliberation outset. The manipulation in *Case 3* is done by rigorous training from the manipulatee's family and community. Given these different means of manipulation, some would say simulation should be put in Place 2 since it is more similar to *Case 2* (as preprogramming is present in both cases) than to *Case 1*.

I think it might be true with respect to the indirectness of the influence a simulator has on the simulatee. However, I think there are other aspects in which simulation is more similar to *Case 1* than to *Case 2*. One such aspect is the power imbalance. The power imbalance in *Case 1* seems to be greater in the sense that the manipulator in *Case 1* can directly invasively violate the integrity of the manipulatee's body/brain while the manipulator in *Case 2* cannot. And the power imbalance in simulation, I argue, is even greater than that in *Case 1*. The simulator can, by simply adding or revising several lines of code, in principle directly invasively violate the integrity of the simulatee's body (though he may choose not to) and preprogram his deliberation outset (as suggested by *the key structure*) in the same time. That said, the manipulation techniques a simulator can employ are arguably the combination of techniques present in both *Cases 1 and 2*, which makes a simulator more powerful and God-like than a manipulator. In light of this, I argue it is not obviously true that simulation is more similar to *Case 1* than to *Case 2* and should be put in Place 2 rather than Place 1.

At the very least, even if simulation in fact resembles *Case 2* more, it should not end up in Place 2 for the following reason. Regarding why simulation does not undermine moral responsibility (to be discussed in § 3), my resource is the discussion of virtual reality which, as far as I am concerned, no moral responsibility specialist has particularly appealed to. That is, my later argument will invite hard-liner hard determinists to a new dialectical environment. If they are on board with my judgment about moral responsibility, if any, in virtual reality, and provided the isomorphism between simulation and manipulation (as I defended earlier in this section), hard-liner hard determinists will be pushed to introspect the intuition they appeal to. To this extent, the introduction of simulation should initiate a new thread, rather than merely adding a new dot to the old thread of the four-case transition.

[21]Or in terms of entailment, the conjunction of prior states and laws of nature entails post states. See van Inwagen (1983: 65) for a definition of determinism in this fashion.

[22]Unless something like a God or a creator is introduced, which is nonetheless not entailed by the thesis of determinism *per se*.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

2.2. Dismissing some irrelevant differences

The precedent section argues: There is no MR-relevant difference between simulation and manipulation. I now anticipate several objections in the subsequent section.

One objection says: Though simulators and manipulators affect one's action in similar ways, simulators are outside the simulatee's world while manipulators are inside the manipulatee's world. And this in-/outside distinction might be MR-relevant. But I argue this alleged difference, though seems plausible, is MR-irrelevant. To see how, consider any video game as a toy simulation. The game developer can, from time to time, enter the game world by logging in to her avatar. That said, the boundary of a simulation is not in principle impenetrable, and simulators are not necessarily outside. Meanwhile, manipulators are not necessarily inside. To make an intuitive case, consider the comic world of John Constantine.[23] In that world, the God lives in Heaven, the Devil lives in Hell, and both agree to not enter the Earth and interfere with humans directly. However, the Devil may still manipulatively seduce humans by making a sequence of events that leads a person to, say, commit suicide.

Another objection says: There might be an MR-relevant difference between simulation and manipulation regarding whether a particular manipulative intention is present.[24] For example, in Pereboom's *Case 1*, the manipulation is directed by the particular intention of having Plum kill White. Simulation, by contrast, is usually directed by the intention of making the state of the world evolve in a particular pattern (e.g., a world in which Nazis won WWII). Such an intention does not primarily care about particular agents like Plum and particular actions like killing white. That said, even if simulation is a kind of manipulation that does not undermine moral responsibilities, other kind(s) of manipulation may undermine moral responsibilities as well. I think this line of reasoning might be true, but it is arguably soft-liner in spirit. Recall Mele's Zygote case or Pereboom's *Case 3*. In these two cases, the manipulator, if at all, does not manipulate the manipulatee by implanting particular intentions like killing White to him, either. However, at least for hard-liners, there is no relevant difference between the Zygote case, *Case 3,* and *Case 1* regarding whether the manipulatee, if at all, should be held morally responsible for his killing White. That said, at least for hard-liners, the presence (or absence) of particular manipulative intention is probably moral responsibility-irrelevant.

A third objection says: Manipulation requires bad intentions while simulation doesn't; therefore, a simulator is not necessarily a manipulator. My reply is this. Sure, the simulator's intention needs not to be bad, but it is arguably irrelevant to whether manipulation in fact takes place. For example, consider the debate over God, free will, and manipulation in scholastic philosophy. A Catholic hard determinist who defends the simulation hypothesis (or its ancient relative, say, external world skepticism) can consistently believe the following: The God doesn't intend to play us around as his puppets (because of his love for us); he has the ultimate control of everything happening in the world; and the world he creates is a simulation (or a dream). In this case, it seems plausible for the Catholic hard determinist to say we are manipulated by God, even though there is no bad intention involved. Besides, recall Pereboom's *Case 1*. In that case, it is true that these scientists manipulate Plum, but they may do so for good, paternalist reasons (like to avoid Plum's suffering), though their "good" intention arguably does

[23]Which is similar to Pereboom's *Case 2*, in which these scientists cannot control what is going on in Plum's brain directly, but they can still manipulate Plum by programming him from the outset of his reasons for action.

[24]My thanks to David Chalmers for pressing this challenge to me.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

not change the fact that Plum is manipulated. To sum up, the manipulator's intention, no matter good or bad, is moral responsibility-irrelevant.[25]

3. Moral responsibility in virtual reality

§ 3 justifies premise 2 of my master argument, that "Even if we are in a simulation, we still have moral responsibilities." The supporting reasons for this premise, I believe, come from both the simulator's (§ 3.1) and the simulatee's sides (§ 3.2).[26] Notably, either reason may independently justify premise 2 of my master argument (though they jointly persuade better).

3.1. The simulator-centric reason

The simulator-centric reason says: A good simulation, from the simulators' perspective, should be designed to allow its inhabitants to be morally responsible because moral responsibility is usually connected to the meaning in life and the meaning in life motivates simulatees to help realize the simulator's goal. Expectedly, this reason is committed to a *moral responsibility-meaning in life* connection, which holds that moral responsibility matters deeply to "our conception as deliberative agents," to whom the absence of healthy "reactive attitudes essential to good human interpersonal relationships and meaning in life" would be disastrous deprivation of life hopes (i.e., aspirations for achievement) (McKenna and Pereboom 2014: 262–79). The *moral responsibility-meaning in life* connection is canonical in the relevant debate. However, it has recently been challenged by, most notably, some hard determinists.[27] They object: There might be a sense in which an agent can satisfy her life hopes or achievements without the presence of moral responsibility, though this sense is probably not as robust as we might naturally suppose. The picture of a meaningful life without moral responsibility can be drawn in three ways.[28] My goal here, then, is to show how these alternatives are, though arguably making sense, probably not the best options for simulators.

One way of accommodating the meaning in life in a world without moral responsibility is to connect the meaning in life to purely consequentialist considerations.[29] That said, one can still live a meaningful life without moral responsibility by committing herself to some purely consequentialist doctrines, e.g., to always maximize her welfare. I understand this idea may seem attractive, especially for those who have the platitude that sims are just robot-like followers of codes. However, I mark two concerns about the possible insufficiency of a purely consequentialist base for meaning in life from a simulator's perspective. Firstly, merely consequentialist doctrines may be incompatible with some

[25]In fact, Pereboom (2001, 2014) does not even include these scientists' intention when making his cases (which I think he is right to do so).

[26]Undoubtedly, as one anonymous reviewer mentions, people with preoccupying philosophical commitments (e.g., the sympathy for hard determinism) will probably not find these arguments as persuasive as I may assume. I think they are reasonable to have conservations. Presumably, there could be another piece about "Are there moral responsibilities in VR?" (which, as far as I can see, does not exist yet). And better arguments for and against the position will be found therein. Nonetheless, the present piece shall at least serve as a friendly invitation for discussion. (In fact, though I explored how simulation may help defend compatibilism in this paper, I think hard determinists may also do something with simulation in defense of their position. And I would be happy to see some pieces about that in the future).

[27]The representative challenger is of course Derk Pereboom, whose arguments my defense of the *moral responsibility-meaning in life* connection primarily targets.

[28]My thanks to Xiaofei Liu for pushing me to consider these objections.

[29]See, e.g., Pereboom (2014: 172–3).

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

simulations. Consider, for example, a trivial case – the simulation of a world in which all agents are virtuous moral saints. For this simulation, arguably, some virtue ethicist principles are also needed for its simulatees to live a meaningful life. More realistically, consider a simulation of the actual world at present: Since not everybody is (and in fact many of us are not) stubborn consequentialist, to simulate our moral deliberation and action, something over and above purely consequentialist considerations seems to be required. Secondly and more crucially, merely consequentialist doctrines may fail to approximate the complexity that some simulation requires. Consider, for example, a simulation of how a certain policy affects the global society. If everybody acts solely in accordance with some purely consequentialist doctrines, then, seemingly, their actions are highly predictable (i.e., they always act to maximize welfare). However, one important goal of the simulator running this simulation is to properly approximate the complexity and unpredictability of people under the given circumstance. That said, at least for this simulation and the like, guiding sims with purely consequentialist doctrines undermines the simulator's goal.

Another way of accommodating the meaning in life in a world without moral responsibility is to connect the meaning in life to some objective attitudes.[30] Some would say that apart from reactive attitudes like resentment or appreciation we have toward our fellow agents that are closely connected to moral responsibility, some objective attitudes we have toward non-agential entities like the appreciation of nature or art may also constitute the meaning in life. I agree with this point. However, I think this point only applies to a small number of simulations (e.g., a simulation of a world governed by an artist king). But I doubt if objective attitudes like the appreciation of nature or art alone are enough, for example, to motivate sims in a simulation of WWII. It seems hard for me to imagine a sim Jewish French patriot devotes herself to the resistance movement merely because of some objective attitudes like "(sim) Hitler launched a war and genocide, though he is not responsible for it in the genuine sense and I do not resent him."

A third way of accommodating the meaning in life in a world without moral responsibility is to connect the meaning in life to some religious commitments.[31] That said, by detaching oneself from taking credit for her deeds, one's life could still be deemed meaningful in the Stoic sense (for her peace of mind), the Christian sense (for her humility), or the like. I think this point might be true. But my concern, for starters, is that religious commitments might be incompatible with some simulations, consider, for example, a simulation of a world in which everybody is an atheist philosopher. Besides, a more serious problem is this: Arguably, a simulator deliberately converting a population of sims to a certain religious commitment is, by analogy to a cult's brainwashing, morally questionable itself, even if not straightforwardly wrong.

Summing up my point, I conclude: At any event, moral responsibility still seems to be the most unproblematic base for a simulator to ground the meaning in life for her simulatees, as most people conventionally assume.

Finally, one objection to the simulator-centric reason may run like this: It might stand for benevolent simulators; however, what if the simulator is malevolent? Then why would he even want to create a good simulation that allows its inhabitants to live meaningful lives? I don't have a strong position against this objection. But I think it is practically unwise, given the cost-benefit calculus, to create simulations that forbid the meaning in life, even for malevolent simulators (like an evil Demon). In practice, simulators use energies and resources (i.e., the cost) to create simulations that serve their

[30]See, e.g., Pereboom (2014: 178–93).

[31]See, e.g., Pereboom (2014: 175, 196–9).

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

purpose (i.e., the benefit). Creating simulations will use a lot of energy and resources, that said, the cost is not cheap (Bostrom 2003: 245). Given such a cost, the simulator will presumably expect reasonable benefits. That said, simulators would naturally prefer their purposes – no matter what, even if it is something narcissistic or perverse like *to be worshiped* or *to play God* – to be achieved, which requires some actions to be done by simulatees (even if it is something like praying).[32] Having the meaning in life by being morally responsible – that is, taking credit for one's achievements in the genuine sense – can probably better motivate actions (including but not limited to simulatee's). Consider, for illustration, a person finds the cure for cancer. What she has done, in *Scenario one*, may be deemed *her* accomplishment (in a genuine sense) and people praise and congratulate *her* for that, or it may not be deemed *her* accomplishment (in a genuine sense) in *Scenario two* and people's praise and congratulation to "her," if any, feels exactly the same with their praise and congratulation to a random stranger. Intuitively, the person will have a stronger motivation to work hard in *Scenario one*, and this intuition in principle generalizes.[33] In light of this, from a practical perspective, in most cases, it seems preferable to create simulations that allow the inhabitants to live meaningful lives by being morally responsible.

3.2. The simulatee-centric reason

There are two classes of standards, as I see it, to judge whether there are moral responsibilities in simulations. One class of standards is internal, which focuses on some internal qualities of (possible) moral responsibility-holders. However, appealing to these internal conditions is probably bootstrapping because, given manipulation arguments, there are concerns about whether usual (compatibilist) internal conditions for MR-susceptibility – e.g., the ability to identify and reflect on desires, the ability to respond to (both moral and non-moral) reasons, and the ability to not act on compulsive or compelled desires, among others – are enough.[34] That said, by looking into manipulation arguments, compatibilists (especially soft-liner compatibilists) intend to figure out whether MR-susceptibility further requires a certain "no manipulation" condition: The efficiency of available internal conditions itself is being questioned.

A better option, then, seems to be turning to the other class of standards, i.e., some key external conditions of the surrounding environment in which agents are held morally responsible for their actions. In § 3.2, I will examine three such external conditions. They are: 1) the presence of persons/agents, 2) the presence of events/actions, and 3) the presence of a certain inter-personal, social structure. These key

[32]There could be some exceptions, though. For example, it is possible that an evil demon just wants to have an idle and lazy simulated population that yawn (or more radically, doing nothing) all day long; after all, he is evil and his preference is probably deviant. But I am not sure why any reasonable simulator would want that – though who knows, some find it unreasonable to create simulations, no matter what! (A more radical stance on this issue is, for example, it is reasonable to *not* create any simulation for the consideration of existential risks, see Greene 2020).

[33]There is a huge body of literature in psychology about human motivations. See Ryan (2012) (especially chs. 6, 12 about self-determination) for relevant discussions (and my thanks to Peter Finocchiaro for drawing my attention to this body of literature). The argument here need not to go that far. All I need is just this: Were there "credits" to be taken from one's action, other things being equal, she would be more motivated. This idea seems to naturally follow from the *moral responsibility-meaning in life* connection justified earlier.

[34]For these desire-based conditions, see, e.g., Frankfurt (1971), and Mele (1995: ch. 4); for the reason-responsiveness condition, see, e.g., Wolf (1987), and Fischer and Ravizza (1998: chs. 2–3).

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

external conditions seem natural. Just consider our standard practice of holding others morally responsible: We usually hold *an agent* (clause 1) of *some moral community* (clause 3) morally responsible for *an action* (clause 2). The question, then, is: Can a simulation satisfy these conditions? My answer is yes. A simulation can satisfy the first two conditions. I report that there seem to be good reasons to believe 1) sims/digital lives deserve moral status and may further bestow autonomy or agency, and 2) virtual events really take place, though I cannot provide a full-scale defense.[35] The third key external condition, that moral responsibilities require a certain inter-personal, social structure, probably needs more demonstration. Inspired by Peter Strawson's famous insight (1962) that moral responsibilities are closely tied to our reactive attitudes, the third condition captures something deep in our conception of moral responsibility. That is: When one person is held morally responsible, other members of her society are also put in the position to resent or appreciate. I argue that there is indeed a certain inter-personal, social structure in a simulation such that a sim can have reactive attitudes toward other sims. Imagine a simulation of human history from 1900 to 2020. In that simulation, a sim mother Terressa can – at least we can interpret her actions as – resent sim Hitler for launching sim genocide. Similarly, a sim member of sim Nobel Prize Committee can also – at least we can interpret their actions as – appreciate sim mother Terressa for her philanthropy.[36]

I now propose this argument: Both virtual reality and physical reality satisfy these key external conditions of MR-susceptibility; there are moral responsibilities in physical reality (at least hypothetically); therefore, if there are moral responsibilities in physical reality, it is highly likely that there are moral responsibilities in virtual reality, too.

Three remarks are in order. First, the second premise is a relatively weak claim. For realists about moral responsibilities, it can be strengthened to be "There are in fact moral responsibilities in physical reality." Since I only intend to argue for a conditional conclusion here, the weak version suffices.

Second, there might be a problem if I argue for a stronger conclusion that "If there are moral responsibilities in physical reality, there are moral responsibilities in virtual reality, too." Since these three key external conditions are best understood as (strong) indicators, rather than as sufficient conditions, of the existence of moral responsibility, the presence of these external conditions does not guarantee the presence of moral responsibility. Hence, the strong conclusion does not follow from my premises, though the weaker conclusion I proposed shall not face this problem.

Third, I clarify "physical reality" in this argument does not necessarily mean the actual world. One potential objection says: Were the "physical reality" here to mean the actual world, this argument dubiously begs the questions against hard-liner hard determinists (take Pereboom for example).[37] Pereboom's position, charitably put, should be something like "Determinism is true for the actual world in which there is no free will (and moral responsibility), regardless of the world being physical or virtual." In this sense, Pereboom would probably reject this premise in the first place.

[35]The primary source of relevant discussion is of course Chalmers (2022). *Re* the moral status of sims, see Chalmers (2022: 331–49); *re* the occurrence of virtual events, see Chalmers (2022: 114–7).

[36]This interpretation, though, invites the debate over whether sims can have attitudes or mentality in general. My tentative position is this: Sims with simulated mind can have consciousness (at least the simulated equivalent of consciousness), see Chalmers (2022: 274–93) for relevant discussions. Anyway, sims are at least behaviorally conscious. As for whether the alleged "consciousness" sims have is the kind of consciousness we think we have (or want to have), see my discussion in § 4.

[37]Here I only discuss Pereboom's position as a representative, which is presumably the most influential. My thanks to Xiaofei Liu for raising this concern.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

I agree it is a legitimate concern. My reply is this. Imagine three possible worlds: Wa, Wv, and Wp. Wa is the actual world. And provided the relevant disagreement among philosophers, let us suspend judgment on whether there is free will or moral responsibility in Wa for now. Wv is a virtual possible world, which is very similar to Wa. That said, the only difference between Wa and Wv is that, from a God perspective, Wa is essentially a host of particles while Wv is essentially a sequence of digits. But for any person in Wa and her counterpart in Wv, their first-person experiences are indiscernible. Wp is a physical possible world that is similar to Wa in all aspects except that there is free will (and moral responsibility) in Wp.[38] Suppose Wv satisfies these key external conditions of MR-susceptibility. Consider, then, the following adaption of my argument: (P1) Wa, Wv, and Wp all satisfy these key external conditions of MR-susceptibility; (P2) there are moral responsibilities in Wp; therefore, if there are moral responsibilities in Wp, it is highly likely that there are moral responsibilities in Wa and Wv, too.

This adaptation, as far as I can tell, does not beg the question against Pereboom. Instead, it may serve as an argument showing why Pereboom is probably wrong about Wa, i.e., the actual world. At any rate, I think Pereboom is disposed to some further issues, even if he rejects the conclusion in the end. Firstly, to reject P1, Pereboom may have to tease out the relevant difference among Wa, Wv, and Wp. But I doubt if it is a viable option since the way I introduce these three worlds leaves little space for further informative, non-*ad hoc* ramifications. Secondly, to reject P2 by questioning the intelligibility of Wp, Pereboom may have to strengthen his position by arguing "If determinism is true, it is necessarily true" and Wp is unintelligible in the first place. But I am not sure whether Pereboom will be on board with this idea (cf. fn. 38). Thirdly, to dismiss the strength of the conclusion, Pereboom may have to argue "very likely" does not guarantee truth, and Wa, Wv, and Wp may still be different in other MR-relevant ways despite their mutual satisfaction of my three key external conditions. I think it is Pereboom's most promising way-out. However, I doubt how, at the end of day, any such alleged "other MR-relevant way" would not turn out to be a "no manipulation" condition that begs the question against me in the first place.

4. Conclusion

In precedent sections, I have argued for the following. There is no MR-relevant difference between simulation and manipulation (§ 2). There are (genuine) moral responsibilities in simulations (§ 3). Provided these two points, and the simulation hypothesis that we are and always have been in an artificially designed simulation of a world, the conclusion is this: Manipulation is compatible with moral responsibilities.

Before closing, I consider one major objection to my position. It says: Since we don't know if there is a simulator external to our reality whose influence on our actions is no different from manipulation (as I argued in § 2), we might as well be led to a sort of skepticism about moral responsibility: It is hard to say whether anybody should ever be held morally responsible for any action at all. I argue this objection is dubiously question-begging. The dialectic here is something like this. When I contest the intuition that "Manipulation undermines moral responsibility," I contest the platitude that

[38]I think Pereboom can probably accept this. As far as I can see, Pereboom has never argued that if his hard-determinism is true, it is necessarily true (viz. true for all possible worlds, on a Leibnizian conception of necessity). All he has argued for is how and why hard-determinism is true for the actual world. To this extent, I think he can accept the possibility that there are possible worlds that are really similar to ours in which compatibilism is nonetheless true.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

MR-susceptibility entails some sort of no-manipulator condition. To get the moral responsibility-skeptic conclusion of the targeted objection, one needs to presuppose something like "If there is a manipulator (simulator) external to our reality that intervenes with our actions, then nobody should ever be held morally responsible for any action at all." But this presupposition begs the question that I contest.

What's more, there seems to be a more dangerous assumption behind this objection. It says: For moral responsibility (in lower case) in virtual reality to be valuable, it has to be exactly the same as Moral Responsibility (in upper case) that we think we have as non-sims. But this evaluative assumption is ungrounded.[39] Furthermore, the assumption is probably embedded in a more general bias toward virtual reality. That is, many would prefer Free Will/Moral Responsibility (in upper case) to free will/moral responsibility (in lower case) even knowing that, if the simulation hypothesis is true, they can only have the latter. As I understand it, such a bias is something like people's preference to have 10 billion dollars, to be beyond-Einstein smart, or to be immortal. All such wishes are preferable, but unreachable (at least for most people) in the same time. They all seem natural to human psychology. However, it seems strange to say our biased preferences as such disvalue our current situation, however unideal. For example, it seems strange to say our preference for immortality disvalues our mortal life. Similarly, it seems strange to say our preference of Free Will/Moral Responsibility (in upper case) disvalues free will/moral responsibility (in lower case) that we are probably having now if the simulation hypothesis is true. To this extent, disvaluing the life in virtual reality based on our bias is unjustified, though natural it is.[40]

I finish the paper by reiterating its moral. The lesson, I believe, is three-fold. First, this paper may account for the relevance of artificial intelligence and its philosophy by providing a case for how the investigation into new technology (like virtual reality) can help resolve problems in traditional philosophy (like problems about free will). Second, this paper may shed light on the long-locked debate between hard-liner hard determinists and hard-liner compatibilists over manipulation. Third, this paper may offer moral responsibility specialists one reminder. Around manipulation arguments, people's previous focus on what, how and in which circumstance the action is caused is, though reasonably well-motivated, arguably short-sighted. Agents surely matter, causes of action surely matter, and environments surely matter, but the nature of reality matters too. To this extent, this paper serves as a blurb calling for moral responsibility specialists' attention to factors in a broader picture which, as far as I can see in the contemporary literature, are mostly underestimated. Such underestimation arguably gets in the way of a more thorough understanding of free will, moral responsibility, manipulation, and determinism.

Acknowledgments. The author would like to thank David Chalmers, Xiaofei Liu, Peter Finocchiaro, and an anonymous reviewer for helpful conversation. Special gratitude is owed to Dave for inspiration from his works and the "Technophilosophy" class, and to Xiaofei, in whose "Moral Responsibility" seminar ideas defended in the present piece first came into being.

[39]But maybe someone can ground a similar assumption. For instance, Grace Helton (2024) recently argues that simulation undermines some social knowledge that depends on other people's genuine mentality, and the life in simulation is therefore less meaningful. See the last part of Chalmers (2024) for a response.

[40]What's more, one may strengthen my position by arguing: Just like our preference to tremendous wealth, intelligence and longevity is probably irrational, our preference to physical reality over virtual reality might be irrational, too. Here I just stick to the weak version.

https://doi.org/10.1017/epi.2025.8 Published online by Cambridge University Press

References

Barnes E.C. (2013). 'Freedom, Creativity, and Manipulation.' *Noûs* **49**(3), 560–88. https://doi.org/10.1111/nous.12043

Bostrom N. (2003). 'Are We Living in a Simulation?' *The Philosophical Quarterly* **53**, 243–55. https://doi.org/10.1111/1467-9213.00309

Bostrom N. (2009). 'The Simulation Argument: Some Explanations.' *Analysis* **69**, 458–61. https://doi.org/10.1093/analys/anp063

Chalmers D. (2022). *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W. W. Norton.

Chalmers D. (2024). 'The Simulation Hypothesis: Metaphysics, Epistemology, Value.' In U. Kriegel (ed), *Oxford Studies in Philosophy of Mind*, Volume **4**, pp. 498–514. Oxford: Oxford University Press. https://doi.org/10.1093/9780198924159.003.0017

De Marco G. (2023). 'Manipulation, Machine Induction, and Bypassing.' *Philosophical Studies* **180**(2), 487–507. https://doi.org/10.1007/s11098-022-01906-2

Deery O. and Nahmias E. (2017). 'Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood.' *Philosophical Studies* **174**(5), 1255–76. https://doi.org/10.1007/s11098-016-0754-8

Fischer J.M. (2011). 'The Zygote Argument Remixed.' *Analysis* **71**, 267–72. https://doi.org/10.1093/analys/anr008

Fischer J.M. and Ravizza M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511814594

Frankfurt H. (1969). 'Alternate Possibilities and Moral Responsibility.' *The Journal of Philosophy* **66**, 829–39. https://doi.org/10.2307/2023833

Frankfurt H. (1971). 'Freedom of the Will and the Concept of a Person.' *The Journal of Philosophy* **68**(1), 5–20. https://doi.org/10.2307/2024717

Gilbert S. (2023). 'The Wrong of Wrongful Manipulation.' *Philosophy & Public Affairs* **51**(4), 333–72. https://doi.org/10.1111/papa.12247

Greene P. (2020). 'The Termination Risks of Simulation Science.' *Erkenntnis* **85**, 489–509. https://doi.org/10.1007/s10670-018-0037-1

Harris K. (2024). 'The Simulation Argument Reconsidered.' *Analysis* **84**(1), 23–31. https://doi.org/10.1093/analys/anad048

Helton G. (2024). 'The Simulation Hypothesis, Social Knowledge, and a Meaningful Life.' In U. Kriegel (ed), *Oxford Studies in Philosophy of Mind*, Volume **4**, pp. 447–60. Oxford: Oxford University Press. https://doi.org/10.1093/9780198924159.003.0014

Herdova M. (2021). 'The Importance of Being Ernie.' *Thought* **10**(4), 257–63. https://doi.org/10.1002/tht3.503

Jeppsson S. (2020). 'The Agential Perspective: A Hard-Line Reply to the Four-Case Manipulation Argument.' *Philosophical Studies* **177**(7), 1935–51. https://doi.org/10.1007/s11098-019-01292-2

Kane R. (1996). *The Significance of Free Will*. Oxford: Oxford University Press.

Lewis P. (2013). 'The Doomsday Argument and the Simulation Argument.' *Synthese* **190**, 4009–22. https://doi.org/10.1007/s11229-013-0245-9

Liu X. (2022). 'Manipulation and Machine Induction.' *Mind* **131**(522), 535–48. https://doi.org/10.1093/mind/fzaa085

McKenna M. (2008). 'A Hard-Line Reply to Pereboom's Four-case Argument.' *Philosophy and Phenomenological Research* **77**(1), 142–59. https://doi.org/10.1111/j.1933-1592.2008.00179.x

McKenna M. (2014). 'Resisting the Manipulation Argument: A Hard-liner Takes it on the Chin.' *Philosophy and Phenomenological Research* **89**, 467–84. https://doi.org/10.1111/phpr.12076

McKenna M. and Pereboom D. (2014). *Free Will: A Contemporary Introduction*. New York: Routledge.

Mele A. (1995). *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press. https://doi.org/10.1093/0195150430.001.0001

Mele A. (2006). *Free Will and Luck*. New York: Oxford University Press. https://doi.org/10.1093/0195305043.001.0001

Mele A. and Robb D. (1998). 'Rescuing Frankfurt-Style Cases.' *The Philosophical Review* **107**(1), 97–112. https://doi.org/10.2307/2998316

Mickelson K. (2015). 'The Zygote Argument Is Invalid: Now What?' *Philosophical Studies* **172**, 2911–29. https://doi.org/10.1007/s11098-015-0449-6

Noggle R (2022). 'The Ethics of Manipulation.' In E.N. Zalta (ed), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2022/entries/ethics-manipulation. Accessed 2 May 2024.

Pereboom D. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511498824

Pereboom D. (2014). *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685516.001.0001

Ryan R.M. (ed) (2012). *The Oxford Handbook of Human Motivation*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.001.0001

Strawson P. (1962). 'Freedom and Resentment.' *Proceedings of the British Academy* **48**, 187–211.

Tierney H. and Glick D. (2020). 'Desperately Seeking Sourcehood.' *Philosophical Studies* **177**(4), 953–70. https://doi.org/10.1007/s11098-018-1215-3

Usher M. (2018). 'Agency, Teleological Control and Robust Causation.' *Philosophy and Phenomenological Research* **100**(2), 302–24. https://doi.org/10.1111/phpr.12537

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

van Inwagen P. (2008). 'How to Think about the Problem of Free Will.' *The Journal of Ethics* **12**(3–4), 327–41. https://doi.org/10.1007/s10892-008-9038-7

White J. (2016). 'Simulation, Self-Extinction, and Philosophy in the Service of Human Civilization.' *AI & Society* **31**, 171–90. https://doi.org/10.1007/s00146-015-0620-9

Winsberg E. (2022). 'Computer Simulations in Science.' In E.N. Zalta and U. Nodelman (eds), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2022/entries/simulations-science. Accessed 19 June 2024.

Wolf S. (1987). 'Sanity and the Metaphysics of Responsibility.' In F.D. Schoeman (ed), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, pp. 46–62. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511625411.003

Cite this article: Wen Z.H. (2025). "Simulation & Manipulation: What Skepticism (Or Its Modern Variation) Teaches Us About Free Will." *Episteme* 1–16. https://doi.org/10.1017/epi.2025.8