How Manipulation Arguments Mischaracterize Determinism

Paul Torek

Forthcoming in *Philosophical Papers*

* pre-publication version; please cite the published version once it is available *

Abstract I outline a heretofore neglected difference between manipulation scenarios and merely deterministic ones. Plausible scientific determinism does not imply that the relevant prior history of the universe is independent of us, while manipulation does. Owing to sensitive dependence of physical outcomes upon initial conditions, in order to trace a deterministic history, a microphysical level of analysis is required. But on this level physical laws are time-symmetrically deterministic, and causality, conceived asymmetrically, disappears. I then consider a revised scenario to resurrect the threat of manipulation even in the presence of time-symmetry and sensitive dependence upon initial conditions. To do so we posit a manipulator containing all the information of the manipulated and time-symmetrically related to him. The new scenario violates special relativity, but even waiving that objection, the scenario faces a dilemma. I argue that the manipulator either lacks agency enough to manipulate the target, or can integrate thoughts separated by long time spans into a single decision. In the latter case, the 'manipulated agent' disappears into the manipulator, making it the story of just one agent.

Keywords Manipulation; determinism; time-symmetry; causation; identity

1. Introduction

Manipulation arguments have become one of the most important critiques of compatibilism in recent philosophical discussion. Neither free will, nor moral responsibility for actions, is compatible with determinism, according to the arguments. Though there are many variants of the argument form, the basic idea of each is first to identify a scenario in which an agent is thoroughly manipulated or originally designed so as to guarantee the performance of a certain action; this agent is deemed neither free nor morally responsible. Second, it is argued, often with the aid of additional scenarios, that there is no freedom- or responsibility-relevant difference between the manipulation (or design) scenario and

an ordinary action in a deterministic world like (or as like as possible to) ours.[1] Thus, it is concluded that if determinism is true in our world, no one acts freely or morally responsibly.

My aim in this article is to point to a neglected difference between manipulation scenarios and merely deterministic ones, which compatibilists can use to resist the second, no-difference premise of manipulation arguments. Adapting arguments by Jenann Ismael (2016) and Carl Hoefer (2002), I explain why determinism does not imply a fixed (agent-independent) path through personal history, while manipulation does. Thus manipulation distinctively impinges on freedom, and may threaten the control of the victims over their own actions.

In section 2, I outline Mele's (2006) original-design argument, and part of the response by Deery and Nahmias (2017), who use an interventionist account of causation to highlight a persistent difference between manipulation and ordinary causation. Bringing some philosophy of science to bear on the issue does help, but we need to dig deeper. In section 3, I argue that the Designer needs detailed knowledge of microphysical states in order to exercise sufficient control over her target. Once we describe the universe in microphysical terms, however, causality, conceived as an asymmetric relation, *disappears* in a universe governed by the *bidirectionally* (in time) deterministic fundamental laws that are entertained in modern physics. We redraw the counterfactual diagrams representing divine intervention of the Designer versus ordinary determinism, and show why the former fixes the agent's fate while the latter does not. In section 4, we consider a revision of the original-design argument to include a bidirectionally deterministic physics, and begin to ponder how strange the new scenario's laws and Designer would need to be. In section 5, I argue that even if we could conceive of such a new Designer, she would either lack requisite knowledge and control, or her relationship to the manipulated human-like organism would be that of person to sub-personal organ, introducing a glaring new relevant

1 Pereboom (2001: 116) claims that the *best explanation* of intuitions about a manipulated agent 'is that his action results from a deterministic causal process that traces back to factors beyond his control.' This might be weaker than a no-relevant-difference claim. I will argue that causality of the required sort need not apply in a deterministic universe.

difference. Section 6 concludes.

2. Mele's argument and an interventionist reply

Alfred Mele's Zygote Argument centers on a thought experiment about Diana, a powerful goddess, and Ernie, whom she designs to fulfill a very specific destiny:

> Diana creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. 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 located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E. […] Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. So he satisfies a version of my proposed compatibilist sufficient conditions for having freely A-ed. (Mele 2006: 188)

Let A be the theft of a certain wallet (name this result Theft). Diana knows that Ernie will be self-controlled, will rationally deliberate before stealing, and so on, and while these may not be intrinsic to her ultimate goal, they are intended results. Indeed Diana must have extremely detailed and complete knowledge of Ernie's life, and control over many aspects of it, in order to control the act of theft by creating the zygote.

The comparison case for the Diana/Ernie scenario is Bernie. Bernie is conceived in the ordinary way, and becomes a competent agent meeting Mele's compatibilist conditions for free agency, or

various other compatibilist conditions. Bernie deliberates just like Ernie, minus the bizarre causal history, and steals a wallet.

Oisín Deery and Eddy Nahmias use an interventionist causal analysis of the two scenarios to resist the claim that manipulated agents like Ernie are relevantly similar to ordinary agents like Bernie. Rather than try to summarize their analysis, I will focus on one important principle. First, they note (2017: 1261; see also Pearl (2000: 12-13)) that causal diagrams and associated structural equations are to be read asymmetrically: the effect is dependent on the cause. Now consider two variables X and W which each exert causal influence over variable Y in background conditions C, either directly, or via one or more chains of intermediaries. Name the $X \rightarrow \ldots \rightarrow Y$ relation R1, and the $W \rightarrow \ldots \rightarrow Y$ relation R2. Other things being equal, R1 is a *stronger causal invariance relation* than R2 if:

> *Stability*: R1 predicts the value of Y across a wider range of relevant changes to the values of C than R2 does. (Deery and Nahmias 2017: 1262–1263)

Stability is the key to Deery and Nahmias's argument for a systematic relevant difference between manipulation and mere determination. Deery and Nahmias discuss another criterion of causal invariance strength, Reliability, but we will not say much about it here.[2]

Deery and Nahmias (2017: 1263–1264) use the concept of equifinality to highlight the difference:

> As Lombrozo (2010: 309–10) puts it, the deliberative activity of intentional agents … exhibits, to a strong degree, *equifinality*—it results in a particular outcome (the intended one) across many different conditions … .

In the Zygote Argument, equifinality shows up in the insensitivity of Ernie's outcome to environmental conditions. Any chance circumstance that might otherwise divert Ernie away from the wallet will be foreseen by Diana, and she will use her goddess powers to ensure that Ernie is not diverted. Deery and

2 The name Stability follows Woodward (2007: 76–77) and Tierney and Glick (2020: 959); likewise for Reliability.

Nahmias (2017: 1266) use the example of an exciting call from a sister that distracts the agents from the wallet. Ordinary agent Bernie would receive the call and be distracted. But Ernie would simply be manipulated to turn off his phone earlier in the day, because Diana anticipated the sister's phone call and prevented it. Diana's decision process predicts the value of Theft across a wider range of background conditions than Ernie's decision process does. Diana's relation to Theft is more stable, and no less reliable than Ernie's,[3] so altogether she bears a stronger causal invariance relation.

The strongest causal invariance relation leading to action in the Zygote scenario is:

(Fig. 1) Diana → Zygote → Ernie → Theft

while that in the Bernie scenario is:

(Fig. 2) Bernie → Theft

Deery and Nahmias (2017: 1267) propose this principle: that an agent can have full freedom and moral responsibility over an action only if her agential structure bears the *strongest* causal invariance relation to the action, among its actual causes.[4] Failing that, the agent may have reduced freedom and responsibility or none; Deery and Nahmias do not attempt to suggest a comprehensive rule. Still, only Diana could be fully free and morally responsible in the Zygote scenario, while Bernie could be so in the ordinary determinism scenario.

Deery and Nahmias mention a second criterion of causal invariance strength, Reliability. Tierney and Glick (2020: 961) argue that two criteria can come into conflict, and that there is no intuitively acceptable way to allocate responsibility while using the criteria to refute the Zygote Argument. And as they point out in a footnote (2020: 967, fn. 25), the requirement that the agent bear the *strongest* causal invariance relation to the result implies a sort of competition for moral

3 Diana is a clockwork goddess, preferring to set things up and let them run. If she were frequently supplying divine interventions to keep things on track, she would be a *more* reliable cause of Theft than Ernie is, as Deery and Nahmias point out in a footnote (2017: 1265). Because other manipulation arguments typically include such meddling, I have chosen the Zygote Argument for its avoidance of this line of objection.

4 Deery and Nahmias (2017: 1263) call the strongest-invariance cause '*the* causal source' (emphasis added). I object to the definite article: superlative sourcehood is not singular sourcehood. So I omit the term from the main text.

responsibility, Intuitively however, moral responsibility typically multiplies when multiple agents act in concert toward a goal; it does not divide, and it is not winner-take-all. I do not propose to evaluate this dispute however, but to examine another feature loosely related to causal invariance strength instead.

Deery and Nahmias focus on differences in strength between various relationships which are all causal. I suggest that we need to examine differences and similarities in the strength of nomological relationships, due to which some of them *fail to be causal* on the interventionist account.

There is another peculiar feature of the Zygote scenario which seems relevant to freedom and moral responsibility: Diana's control over Theft is both very strong and *independent of the specifics of Ernie's deliberations*. We might say that Diana's control *usurps* Ernie's otherwise-apparent control. In championing this point, I am close to agreeing with Pereboom's (2001: 116) claim that the best explanation of a manipulated agent's unfreedom or non-responsibility 'is that his action results from a deterministic causal process that traces back to factors beyond his control.' Isn't this a feature that ordinary determinism *shares* with Manipulation cases? No, but we need to be careful about the definition of 'determinism' - careful to hew close to scientifically viable theories and interpretations. To that end, the definition used by Deery and Nahmias (2017: 1257) is too strong: 'for each 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.' Recall that causes, and the arrows in causal diagrams, are understood *asymmetrically*.

For reasons to be explained, any process in a mundane world going back 30 years that deterministically leads to Bernie's committing Theft must be described in microphysical terms. But at this level of description, it is not a *causal* process because the required asymmetry is lacking.

3. Universal determinism does not imply universal causality

The best accounts of the fundamental laws of physics of our actual universe, general relativity and the Schrödinger equation of quantum mechanics, are Charge-Parity-and-Time (CPT)-reversible.[5] The fact that there is such a simple mapping between forward- and backward-evolving dynamic equations is remarkable, but all we need to note is one consequence of this: from a description of present microphysical states, we can equally well derive past states or future states. Relativity and the Schrödinger equation fit the entailment definition of determinism: a description of the universe at a time, in conjunction with the laws of nature, entails descriptions of the universe at (at least) later times [van Inwagen (1983), Carroll (2010: 108, fn 111)]. Time-symmetric determinism implies that information is conserved, in the sense that the state of the universe at any time nomologically implies the state at all other times [Carroll (2010: 140)]. It also implies that different states at a given time would have evolved from, and would evolve into, different states in the past and future.

Another feature of physics as we know it is *chaos*. The best-known example is weather. Lorenz (1993) pointed out that nonlinear dynamics make our best models of the weather extremely sensitive to tiny variations in initial conditions - the so-called butterfly effect. In the years since, similar sensitivity has been hypothesized to apply to a wide variety of physical systems. Such sensitive dependence on initial conditions has also been found in some studies of the brain [Canavier and Shepard (2009)]. Microscopic differences in initial conditions can easily ramify into macroscopic differences later in such systems. Thus, to guarantee that Ernie commits Theft, Diana needs to know the entire microphysical state of the universe, at least out to the radius (30 light-years) of physical information capable of influencing Ernie's action.

But as long as we are considering microphysical descriptions of large chunks of the universe, why not go for broke? Let's model the entire universe from the time (in our conventional coordinate

5 There are interpretations and theories which are not (even in an extended sense) time-symmetric, such as Ghirardi Rimini and Weber (1986) and Unger and Smolin (2014), but the best-known of these are indeterministic.

system) of the zygote through the time of the theft. To remind us that certain variables now represent universe-wide microphysically detailed states, let us write them in all capitals. Then for the ordinary deterministic case we have

(Fig. 3) … ZYGOTE ↔ BERNIE ↔ THEFT

while for the Diana scenario we have

ENVIRONMENT

↓

(Fig. 4) Diana → ZYGOTE ↔ ERNIE ↔ THEFT

The double-headed arrows of Fig. 3 reflect the bidirectional nature of microphysical determinism. At the microphysical level, not only is there sensitive dependence on initial conditions, but also sensitive dependence on final conditions. But the divine Diana gets a single-headed arrow. This is an aspect of a thought-experiment worth trying: let the goddess be non-physical.[6]

Since Diana is not physical, we cannot assume that laws which may apply to her are bidirectional. More importantly, single-headed arrows express the way we ordinarily think about agents' control over their effects on the world, and we want to faithfully represent the features that drive our intuitions. At human size-scales, conditions at later times are sensitively dependent on earlier times, but not vice versa. (Since Diana acts upon ZYGOTE to achieve an otherwise unlikely result, the value of ZYGOTE cannot simply correspond to the value of ENVIRONMENT, and so the arrow from ENVIRONMENT to ZYGOTE is also unidirectional.)

We have been treating causation as an antisymmetric relation: if A causes B, B does not cause A. The double-headed arrows exclude such causation, and Figure 3 is not a *causal* diagram. To coin a term to apply to nomological relationships regardless of their symmetry or asymmetry, let us say that such variables *influence** each other. In the merely deterministic scenario, ZYGOTE is *not*

6 Is it conceivable that a non-physical being affects the physical world? Yes, provided each realm has its own tightly knit rules or laws, with only sparse interactions. But if the reader doubts it, not to worry: we consider a physical Diana next.

independent of Bernie, since all of BERNIE (the universe-wide state) including Bernie (the human being) is nomologically relevant to (influences*) ZYGOTE. This seems bizarre. We intuitively think of the past as being *there anyway*, as Jenann Ismael (2016: 149) aptly puts it, no matter what we do. But no wonder, given that we are macroscopic beings. Everyday experience does not generalize to the micro-scale. Any change in Bernie, if evolved backward in time, would (by conservation of information) result in a different state of ZYGOTE. Similar points would apply to still earlier states of the universe if we bothered to consider them.

Influence* is not causation; Bernie doesn't *cause* ZYGOTE to be the way it is. Could a symmetric relationship ever explain away an apparent obstacle to freedom? Yes.

Consider a bidirectional nomological connection between a present action and a far future event. Suppose that, in the actual world, Alice leaves for work at 8 am, stirring some breezes by her front door, and at 8 am two weeks later, a certain snowflake blows off the Antarctic Larsen C ice shelf. Suppose that in various very close possible worlds with the same laws, Alice leaves for work at 7:59:59 am or 8:00:01 am (etc.), and at 8 am two weeks later on the Larsen C ice shelf, various different snowflakes blow away while the aforementioned snowflake stays put. Yet Alice lacks any knowledge of the mappings between her actions and Antarctic snow movements. So she has no control whatsoever over how snowflakes in Antarctica are influenced* by her actions. Along comes a philosopher and proclaims that, since her action results from a deterministic process that traces *forward* to factors beyond her control, Alice has no control over when she leaves for work. How plausible is that argument? Yet this is exactly the mirror image of reaching into the past for deterministic antecedents of action.

It might be objected that we should not be so quick to embrace our conviction that forward-going nomological correlations are innocent, and overturn our intuitive sense that backward-tracing nomological relationships threaten control. Why not the reverse? Because to curtail our freedom and

control, an obstacle must be *there anyway*, regardless of what we do. We are accustomed to assuming that the past is there anyway, because the familiar macroscopic past is indeed so. But now that we have learned that we were overgeneralizing, and that the microscopically detailed picture works differently, we should revise our thoughts about those backward-tracing connections.

Let us return to comparing the bidirectional determinism of Figure 3 to the original scenario of Figure 4. By contrast to the ZYGOTE state of Figure 3, Diana stands firm. No influence* from Ernie can touch her. Nor could Ernie's deliberations, had they gone differently, keep her from securing her intended result of Theft: there are many other motives, other than Ernie's actual ones, which Diana could use to effect Theft. In this way, Diana usurps Ernie's would-be control. This is a difference that intuitively makes a difference, to Ernie's lack of freedom, control, and perhaps also moral responsibility.

I should note that, although I intend to make more room for compatibilists to resist the 'no-difference' premise of manipulation and original-design arguments, I don't offer an overall strategy for compatibilists. I don't say whether or in what scenarios to resist the other main premise – that the manipulated agent is neither free nor responsible – as well. However, I do suggest that it may be worth picking apart the topics of control, free will (understood in a way deferential to common use rather than stipulating a definition via moral responsibility), and moral responsibility, and drawing different lines.[7] Distinguishing among various dimensions of moral responsibility may also help (see Kapitan (2000)).

4. A physical goddess?

So far we have stuck with Mele's portrayal of Diana as a divine being, exempt from physical laws. But if we can reconstruct the Manipulation Argument to conform to a sort of determinism

7 Drawing different lines for these topics is by no means new. For example, Fischer (2014) denies that originally-designed agents have full free will (they cannot do otherwise) but affirms their moral responsibility in at least some cases. On the divergence of free will from moral responsibility among lay users of these terms, see the data in Spitzley (2015).

resembling what might plausibly obtain in our world, and still draw ominous implications for free will and moral responsibility, we should. Can we repair Mele's argument by making Diana herself part of a natural world that is both sensitively dependent upon initial conditions, and subject to bidirectionally deterministic laws? Call this the Physical Goddess Argument. Consider the diagram:

(Fig. 5) … ↔ DIANA ↔ ZYGOTE ↔ ERNIE ↔ THEFT

If the scenario is possible, and if we still have the intuition that originally-designed Ernie is not free and not responsible, and if we do not generate some new relevant difference between the design case and ordinary actions in a deterministic world, then a stronger Manipulation Argument is available. However, there are reasons to doubt that the scenario is possible. And the new scenario does generate a new relevant difference. Let's begin with the possibility problem.

In order to design Ernie and thus the zygote to achieve Theft, Diana needs to know what she is doing. As a physical being, Diana must have physical sub-components whose states represent Ernie, the zygote, and the entire environment capable of influencing Ernie, which would be the entire volume out to 30 light-years from the planned theft. She must also have components to do the necessary calculations using this data. Diana herself would need to be located within a smaller distance from Ernie, such as one light-day if she takes a day to create the zygote. There is not enough physical information in that volume of space to accommodate this second copy of all the information plus additional calculation hardware. Moreover, the information from the environment must be collected, and then a physical signal sent from Diana to control the production of the zygote; but the round trip of information would have to violate special relativity's speed limit on information transfer, the speed of light. For if we allow 30 years for Diana to collect the information and one day for her to calculate and send the control signal to the zygote, then she actually needs to collect from a volume of radius 30.003 (rounded to three decimal places) light-years, contradicting the original assumption.

So at the very least, the proponent of the new Physical Goddess Argument will have to imagine

a universe with different physical laws than ours. This need not be a fatal flaw, since the argument is about responsibility and freedom, and it seems possible to imagine beings a lot like us, in terms of what matters for their responsibility and freedom, who inhabit a universe where special relativity is false. However, not all aspects of our actual physical laws should be up for grabs. The differences in the laws must not all by themselves explain Ernie's lack of freedom. Thus, I suggest, we should leave sensitive dependence on initial conditions, and its consequence, chaos, in place. This is a feature that makes our human-scaled world much more interesting than a cartoon stereotype of determinism, and underlies some phenomena which lead most ordinary people to suspect that determinism is false. Further, let us leave time-symmetric determinism in place, because it simplifies the case against supposing that the detailed past is *there anyway*, independent of what we do,[8]

The proponent of the Physical Goddess Argument needs to argue the new physics neither nullify Diana's control, nor eliminate Ernie's unfreedom. Here the proponent faces a dilemma. The relationship between Diana's thoughts and the relevant events lacks the robustness of asymmetric causal relationships, so that it becomes unclear that she possesses genuine knowledge and control. But if we relax the requirement of robustness, a new problem arises: the Physical Goddess's control threatens to be too perfect, swallowing Ernie's status as a separate agent.

5. Knowledge, control, and agential identity

To see the problem, we first need to understand the deeply weird structure of the Physical Goddess scenario. As mentioned in section 3 above, 'ERNIE' refers not just to Ernie but a wide set of environmental states as well: everything that influences Ernie's decision to steal the wallet, including everything that Diana intended. She intended not just that Ernie steal, but that his decision proceeds from considerations of certain reasons, and so on. Environmental states influencing those aspects of

8 Not that it is necessary to that case; see Ismael (2016: 154-55).

Ernie's decision would include an enormous number of facts about Ernie's social and physical world. However – and this is where it gets weird – 'DIANA' need only refer to Diana, now herself a physical organism but somehow comprehending all the physical information relevant to both Ernie and his environment. Time-symmetrically deterministic laws preserve information, and thus Diana contains *all* the information about ERNIE, however she divides it between things she takes as givens (perhaps some of the structure of Ernie's society) versus things she decides (heritable characteristics designed into the zygote).[9] There cannot be any environmental influences which do not run through Diana, even if such nomological relationships would confirm (overdetermine) Diana's intended result.[10] This is why we didn't need to show 'ENVIRONMENT' in our diagram of the Physical Goddess scenario (Figure 5). All the ZYGOTE and ERNIE information is *integrated into Diana*, and it is sufficient to determine the outcome to satisfy all her intentions.

This raises serious questions about Diana's agency. Does Diana *know* all this physical information, or are we simply saying that Diana *is* all this information? After all, knowledge is robust, capable of responding appropriately to different situations. Physical memory systems as we know them depend on the increase of entropy, and a coarse-grained analysis of the memory states, to achieve robustness [Mlodinow and Brun (2014)]. But coarse-graining means loss of micro-scale information, which Diana cannot afford in a world sensitively dependent upon initial conditions. If Diana lacks knowledge of relevant details and fails to exercise sufficient agency to control Ernie's deliberations, the Physical Goddess thought experiment ends here. In my view, it does end here. Knowledge does require the sort of robustness that coarse-grained relationships provide. But we should temporarily waive this objection, and explore the alternative.

9 As Markus Schlosser (2015: 77) notes, 'we are asked to assume that Diana can control Ernie's entire development and his behavior over a period of 30 years merely by arranging the initial properties of Ernie's zygote in a certain way. This, it should be noted, is truly and utterly incredible.' Indeed. But we have already allowed for new laws of physics.

10 Such overdetermination would violate time-symmetric determinism because then a single present state would be compatible with multiple past states.

So instead, let us count Diana's information-conserving nomological connection to and from all these microscopic details as constituting knowledge and influence sufficient for agency. Note that we wrote influence without the asterisk, because now we are supposing that this counts as influence in the intuitive sense relevant to control, as well as the technical sense given for influence*.

But this, together with the fact that all ERNIE information is integrated into Diana, makes Ernie's decision-making into a proper subset of Diana's, absorbing him into a larger person. Ernie's entire mind is integrated into Diana's thoughts. This is hard for us to grasp because we live in a macroscopic, entropy-increasing world with a well defined arrow of time. In our life-world, it is impossible for thoughts to be integrated over a large time span. We automatically think that earlier thoughts influence later ones but not vice versa. But by the definition of the Physical Goddess scenario, Diana is different. The nomological arrows connecting her to ERNIE are double-headed. So while we cannot say that Diana *causes* Ernie's thoughts in an asymmetric way nor that Ernie's cause hers, we can say that both heavily *influence** - and hence, we are supposing, influence, sans asterisk - each other. And this suffices to bring Ernie within a larger agent, much as a single region of a human brain, even an intelligent region such as a hemisphere, is a component of a larger person.

In actual human beings, exchanges of information between brain regions only go forward in time. For example, an outfielder in baseball faces a challenging line drive. An oversimplified and partial account of her brain activity might be as follows. Her prefrontal cortex entertains diving for the ball, and milliseconds later, her motor cortex sends back the message that the plan would fail. A few more milliseconds, and her prefrontal cortex plans throwing her glove to stop the ball, and immediately thereafter, her motor cortex executes the new plan. (She saves the day, holding the batter to a triple.) None of these signals can go backward in time because all of these brain operations increase entropy as, for example, glucose is metabolized, and in our actual history and at size scales even approaching the brain's, entropy only increases with time.

However, by stipulation, Physical Goddess Diana and her relation to Ernie is different. At the time of zygote design, she has complete detailed knowledge about Ernie's decision processes, and the quasi-causal arrows in the network are symmetric, even across time. Remember that we are counting these symmetric relationships as knowledge-providing, in order to deem Diana sufficiently knowledgeable to maintain control. Diana (30 years prior to Theft) and Ernie share *more* information exchange than two regions of a normal human brain. This is what makes Ernie's ratiocination an integral subset of Diana's, rather than that of a separate agent. If we decide to punish 'Ernie' for Theft, this can be reasonable only if we thereby punish Diana.

6. Conclusion

In an essay titled 'The Cosmic Now,' physicist Anthony Aguirre (2019) explores the violent clash of intuition, which even a physicist can feel, against special relativity's verdict that there is no such thing as absolute simultaneity. The idea that there is any level of description at which mechanisms have no preferred direction of time also fiercely clashes with intuition. But if philosophers wish to explore the implications of scientific determinism, they owe it to themselves to consider the leading deterministic theories. When we do, we find that manipulation scenarios highlight not a commonality between manipulation and determination, but a difference. Manipulators are, while mere nomologically sufficient antecedents are not, *independent* of us.

Macroscopic human agents exert one-way control over later macroscopic variables by converting lower entropy energy forms to higher, and our acquisition of knowledge similarly has a temporal bias due to entropy gradients.[11] We can imagine a different physics in order to try to reconstruct a manipulation argument, but a manipulator sufficiently powerful to get the argument

11 This will have to be merely asserted here. See Ismael (2016), Albert (2015), Kutach (2013), and Mlodinow and Brun (2014) for explanations.

started has a very different form of agency from our familiar kind. Because of this, the manipulator either lacks required knowledge, or swallows the manipulated, the latter becoming a proper part of the powerful agent. Manipulation arguments may work well against intuitive conceptions of determinism, but not so well against scientific ones.

References

Albert, David Z. (2015). *After Physics*. Cambridge: Harvard College.

Aguirre, Anthony (2019). The Cosmic Now, https://aeon.co/essays/is-that-leaf-falling-here-and-now-cosmic-koans-on-time

Canavier, C.C. and P.D. Shepard (2009). Chaotic versus stochastic dynamics: A critical look at the evidence for nonlinear sequence dependent structure in dopamine neurons, *Journal of Neural Transmission Supplementa, 73,* 121–128. https://doi.org/10.1103/PhysRevE.89.052102

Carroll, Sean M. (2010). *From Eternity to Here: the Quest for the Ultimate Theory of Time*. New York: Penguin.

Deery, Oisín and Eddy Nahmias (2017). Defeating Manipulation Arguments: Interventionist causation and compatibilist sourcehood, *Philosophical Studies, 174*, 1255–1276. https://doi.org/10.1007/s11098-016-0754-8

Fischer, John M. (2014). Review of Free Will, Agency, and Meaning in Life, by Derk Pereboom. *Science, Religion & Culture, 1,* 202-208.

Ghirardi, G.C., A. Rimini, and T. Weber (1986). Unified dynamics for microscopic and macroscopic systems, *Physical Review D, 34*, 470.

Hoefer, Carl (2002). Freedom From the Inside Out. In *Time, Reality and Experience* (pp. 201-222). Ed. C. Callender, Cambridge: Cambridge University Press.

Ismael, Jenann (2016). *How Physics Makes Us Free*. New York: Oxford University Press.

Kapitan, Tomis (2000). Autonomy and Manipulated Freedom. *Philosophical Perspectives 14* (2000), 81-104

Kutach, Douglas (2013). *Causation and its Basis in Fundamental Physics.* New York: Oxford University Press.

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions, *Cognitive Psychology 61*(4), 303–332. https://doi.org/10.1016/j.cogpsych.2010.05.002

Lorenz, Edward N. (1993). *The essence of chaos*. Seattle: University of Washington Press.

Mele, Alfred R. (1986). Manipulation, Compatibilism, and Moral Responsibility, *The Journal of Ethics, 12*, 263–286

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

Mlodinow, Leonard and Todd A. Brun (2014). Relation between the psychological and thermodynamic arrows of time, *Physical Review E, 89*, 052102. https://doi.org/10.1103/PhysRevE.89.052102

Pearl, Judea (2000). *Causality: Models, Reasoning, and Inference*, 2nd ed. New York: Cambridge University Press.

Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.

Schlosser, M.E. (2015). Manipulation and the Zygote Argument: Another Reply. *The Journal of Ethics, 19*, 73–84. https://doi.org/10.1007/s10892-014-9183-0

Spitzley, Jay (2015) The Importance of Correctly Explaining Intuitions: Why Pereboom's Four-Case Manipulation Argument is Manipulative. *Journal of Cognition and Neuroethics 3* (1): 363–382.

Tierney, Hannah and David Glick (2020). Desperately Seeking Sourcehood. *Philosophical Studies, 177*, 953–970. https://doi.org/10.1007/s11098-018-1215-3

Unger, Roberto Mangabeira and Lee Smolin (2014). *The Singular Universe and the Reality of Time.*

New York: Cambridge University Press.

Van Inwagen, Peter. (1983). *An Essay on Free Will*, Oxford: Clarendon Press.

Woodward, J. (2007). Causation with a human face. In A. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited* (pp. 66–105). Oxford: Oxford University Press.