Manipulation Arguments and Libertarian Accounts of Free Will

Taylor W. Cyr

Forthcoming in the *Journal of the American Philosophical Association*; please cite published version.

Abstract: In response to the increasingly popular manipulation argument against compatibilism, some have argued that libertarian accounts of free will are vulnerable to parallel manipulation arguments, and thus manipulation is not uniquely problematic for compatibilists. The main aim of this paper is to give this point a more detailed development than it has previously received. Prior attempts to make this point have targeted particular libertarian accounts but cannot be generalized. By contrast, I provide an appropriately modified manipulation that targets all libertarian accounts of freedom and responsibility—an especially tricky task given that libertarian accounts are a motley set. I conclude that if manipulation arguments reveal any theoretical cost then it is one borne by all accounts according to which we are free and responsible, not by compatibilism in particular.

Keywords: manipulation arguments, compatibilism, libertarianism, indeterministic manipulation

1. Introduction

An increasingly popular argument against compatibilism (about causal determinism and the freedom required for moral responsibility) is the manipulation argument.[1] There are several versions of the manipulation argument, but the typical structure is as follows. First, a case is described in which an agent is manipulated to perform an action while satisfying some compatibilist set of sufficient conditions on free and responsible action, and it is claimed that the agent lacks freedom and responsibility for performing the action. Next, it is claimed that there is no freedom-and-responsibility-relevant difference between the manipulated agent and ordinary agents in deterministic worlds. Given the lack of freedom and responsibility of the manipulated agent, and given the no-relevant-difference claim, the argument concludes that ordinary agents in deterministic worlds lack freedom and responsibility, which is to say that compatibilism is false.

[1] The argument is typically focused on an agent's (non-)*responsibility* for an action rather than her *freedom* (whether that be her free will or her freedom in acting, if these are distinct), and so, typically, the conclusion of the argument is about the incompatibility of moral responsibility and determinism. Given that the argument can be adjusted to target the compatibility of freedom (of either type) and determinism, though, I will not limit myself to talking about moral responsibility here but will treat the argument as concerned with this cluster of issues, not just responsibility.

One way to respond to the manipulation argument is to deny that the manipulated agent lacks freedom and responsibility.[2] This has been called the "hard-line" reply to the manipulation argument, and, according to some, this reply requires biting a bullet and thus reveals a theoretical cost that is unique to compatibilism. As I argue in this paper, however, an appropriately modified manipulation argument can be wielded against *libertarian* (incompatibilist) accounts of free will. Now, this is not an altogether novel idea. Kristin Mickelson (2015), for example, argues that manipulation arguments support *impossibilism*, the view that freedom and responsibility are impossible for agents like us.[3] But, unlike Mickelson's, my strategy will focus on distinctively libertarian requirements for freedom and responsibility. Even so, I am not the first to present a manipulation argument against a libertarian account of free will, but, as I will argue in this paper, no other such presentation has been successful (nor do they aim to be comprehensive). Since libertarian accounts are vulnerable to manipulation arguments as well, manipulation arguments do not reveal a cost of accepting compatibilism in particular; rather, if they reveal a cost at all, it is a cost borne by any account according to which we are free and responsible.

In section 2, I discuss one widely discussed version of the manipulation argument against compatibilism and explain why some have taken such manipulation arguments to reveal a unique cost of compatibilism. Then, in section 3, I argue that two recent attempts to raise manipulation-related worries for libertarianism do not succeed. In section 4, I present my own indeterministic manipulation scenario and claim that the conditions on free will proffered by extant libertarian accounts are satisfied by the manipulated agent in that scenario. I go on, in section 5, to provide

[2] See, for example, McKenna (2008) and Fischer (2011).

[3] Interestingly, as an anonymous reviewer points out, my own conclusion is very similar to Mickelson's, since I think that manipulation arguments divide those who think we have freedom and responsibility from the skeptics, not the compatibilists from the incompatibilists.

support for this claim by surveying the conditions proffered by a variety of types of libertarian accounts of free will and by showing that these conditions are indeed satisfied by the manipulated agent in my scenario.

2. Manipulation Arguments and the "Cost" of Compatibilism

The worry that agents might be manipulated into satisfying some alleged sufficient conditions on free and responsible action is not a new worry, but two recent manipulation arguments have brought much attention to the worry.[4] The first of these manipulation arguments is Derk Pereboom's "Four-Case Argument" (2001: 110-117; 2014, chapter 4), and the second is Alfred Mele's "Zygote Argument" (2006: 184-196). For the sake of brevity, we will consider only the latter in detail.

The Zygote Argument begins by presenting its audience with a case in which a goddess wants a certain event to occur and in which she creates an agent, Ernie, who she knows will bring about that event. Here is part of Mele's description of the case:

> 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*. (2006: 188)

[4] For examples of earlier discussions of manipulated agents, see Fischer (1994, chapter 1), Fischer and Ravizza (1998, chapters 7 and 8), Kane (1996, chapter 5), Mele (1995, chapter 9), and Watson (1987; 1999).

Mele goes on to stipulate that Ernie satisfies one version of his own proposed compatibilist sufficient conditions for free action, and it is worth noting that we can stipulate that Ernie satisfies *any* proposed compatibilist sufficient conditions for free and responsible action (provided that those conditions are not tailored to rule out the responsibility of designed agents).[5] Mele then argues as follows:

1. Because of the way his zygote was produced in his deterministic universe, Ernie is not a free agent and is not morally responsible for anything.
2. Concerning free action and moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal human zygote comes to exist in a deterministic universe.
3. So determinism precludes free action and moral responsibility. (2006: 189)

Mele notes that premise 1 is a judgment about a case, and he says, "Premise 1 has some intuitive pull on me, but not enough to move me to accept it. I am agnostic about premise 1, as I am about compatibilism" (2006: 194). If one were to be convinced that Ernie lacked freedom and responsibility in the case described, though, Mele thinks that one should be inclined to accept the conclusion, which is that compatibilism is false.

Focusing mainly on Pereboom's Four-Case Argument (but intending for his reply to address any instance of the general form of the manipulation argument), Michael McKenna (2008) argues that compatibilists should grant that certain manipulated agents can nevertheless be free and responsible.[6] McKenna's "hard-line" response grants that there is no relevant difference between

[5] For examples of the conditions mentioned in that parenthetical clause, see Barnes (2015), Deery and Nahmias (2017), and Waller (2014).

[6] A bit more precisely, McKenna thinks that some instances of manipulation undermine responsibility, even by the compatibilist's own lights, but that, eventually, once the manipulation

ordinary determined agents, on the one hand, and agents like the ones who have been manipulated in the cases we have considered, on the other. But McKenna also argues that, since it would be dialectically infelicitous for the proponent of the argument to begin by asserting that ordinary determined agents are not free, there is nothing to prevent us from generalizing from its not being evident that ordinary determined agents are not free to its not being evident that the manipulated agents are not free. And given the dialectical burden borne by the proponent of the manipulation argument, McKenna argues, this much suffices for the compatibilist to claim victory.

But proponents of manipulation arguments have pointed out that taking the hard-line response is tantamount to "biting the bullet" and that this reveals a unique cost of compatibilism. Patrick Todd, for example, says: "Is it really plausible to think that the fact that Plum [a manipulated agent] got such a raw deal at the hands of the neuroscientists [manipulators] is simply irrelevant to Plum's moral desert? I do not think so, but such a result appears to be the (increased) cost of compatibilism" (2011: 133). Fischer mentions this type of remark in his own hard-line reply to Mele's Zygote Argument: "But it is often said that the Zygote Argument and similar arguments display the 'price' of compatibilism—they indicate what a compatibilist must be prepared to accept, where this is somehow more 'philosophically expensive' than compatibilism was antecedently thought to be" (2011: 271). Fischer goes on to say that he does not see how it could be a "cost" of his view of responsibility that it entails that agents are responsible in cases that are not relevantly different from ordinary ones (in which there is no special reason to call into question an agent's responsibility). As Todd says in response, though, "the mere fact that one's view predicts or entails a certain result does not imply that its having that result is no cost for one's view" (2013: 197).

scenario is sufficiently tweaked, compatibilists will have to admit that manipulated agents can be responsible. Thanks to an anonymous reviewer for suggesting clarification here.

Even McKenna notes that, since "the compatibilist will have a very tough time providing a positive argument for her thesis about manipulation cases...these considerations call attention to the compatibilists' limitations" (2008: 157).

If it were to turn out that worries about manipulation also posed a threat to *libertarian* accounts of free will (which are *in*compatibilist accounts), however, then it would be false that manipulation arguments reveal a unique cost of compatibilism.[7] Later in this paper (section 4), I will argue that cases of manipulation generate a worry for libertarian accounts of free will, too, and thus that compatibilists are not alone in being vulnerable to worries about manipulation.

3. Recent Manipulation-Based Worries for Libertarian Accounts

Before I raise this parallel challenge to libertarianism, though, I will survey two attempts to show that manipulation is a problem for certain libertarian accounts of free will, the first presented by Ishtiyaque Haji and Stefaan Cuypers (2001) and the second by Roger Clarke (2012). Although, as I will argue, neither of these attempts succeeds, a critical discussion of them will highlight the advantages of the indeterministic manipulation scenario that I will introduce in the next section.[8]

[7] As an anonymous reviewer points out, a cost shared is still a cost. While I agree and think that counting such costs is crucial to an overall assessment of the relative advantages and disadvantages of positions in the free will debate, I do not take a stand in this paper on whether we should think that manipulation arguments undermine *all* accounts of the conditions on free and responsible action or that *neither* compatibilist nor incompatibilist conditions are undermined by the arguments.

[8] King (2013) attempts to present a case of indeterministic manipulation in which Kane's (1996) conditions are satisfied, but, as I argue in Cyr (2016), Kane's conditions are not satisfied by the agent in King's case. I also argue, however, that the case can be modified such that the agent does satisfy Kane's conditions, and the modified case allows for a challenge to Kane's brand of libertarianism. Using a different type of case, the present paper extends this challenge to all types of libertarian accounts of free will, but it would also be possible to tweak the case already developed into a case in which other libertarian conditions were satisfied.

Haji and Cuypers present the following case of covert nonconstraining (CNC) manipulation (which means that the manipulated agent is unaware of the manipulation but is not coerced or otherwise constrained by the manipulators):

> Imagine that neurology and neurosurgery have so progressed that not only can particular pro-attitudes like desires, volitions, intentions, or goals be induced in an individual (with or without the individual's consent or knowledge), but where one individual can be molded psychologically to be just the kind of person the surgeon desires. Jenny is an adept painter and a gastronome. Jim, though no connoisseur of food and drink, is an adroit computer hacker, having successfully masterminded several "hacking" offenses. Max, the eccentric neurologist, eager to test a new form of psychosurgery, kidnaps and anesthetizes Jenny, turning the artist into Jim's psychological twin.
>
> Devoid of any suspicion that she has fallen victim to Max, Jenny awakens from her surgery with profound changes which, from her own inner perspective, she can only accept. The psychosurgery has endowed her with a new set of values, goals, preferences and the like, while "erasing" ones she formerly had…Catching the morning news, she learns about the new computing system in a bank in Brussels, and after diligent work, manages to transfer from an account in that bank a large sum of money into her own holdings. "Success!" exclaims Max to himself. (2001: 222)

Since Jenny apparently satisfies some libertarian conditions on free and responsible action, and yet intuitively Jenny is not responsible for her hacking, Haji and Cuypers take themselves to have

shown that manipulation is equally as problematic for libertarianism as it is for compatibilism.[9] But, as Haji and Cuypers note (2001: 228), some libertarian accounts are sophisticated and include historical conditions on free and responsible action that are not satisfied in cases of manipulation like Max's.[10] Haji and Cuypers' response to this objection to their project is to point out that the same tactic for responding to worries about manipulation "could equally well be deployed to rescue compatibilist competitors from the threat of CNC manipulation" (2001: 230). Haji and Cuypers are wrong about this, however, since even compatibilist accounts that include historical conditions—such as Fischer's (2006; 2012) or Mele's (1995; 2006)—entail that the agents in certain manipulation cases (such as Ernie in the case from Mele's Zygote Argument) are morally responsible for what they are manipulated to do. To show that manipulation is a problem for libertarianism as well, what is needed is an indeterministic manipulation scenario in which even the various accounts' historical constraints on free and responsible action (that would preclude manipulation like Max's from resulting in free and responsible action) could be satisfied.

---

[9] In its current form, the case is under-described, leaving it implicit that the manipulated Jenny satisfies typical libertarian conditions. Haji and Cuypers go on to consider a few libertarian accounts of free will—Mele's (1995) event-causal "modest libertarianism," Kane's (1996) event-causal libertarianism, and Clarke's (1995; 1996) agent-causal libertarianism—and to tweak the case (or to replace it) as is necessary to challenge each account. As it turns out, they fail to show that Jenny satisfies Mele's conditions (see note 8), and their sample of libertarian accounts is too small. (It is unclear, for example, whether their worries extend to non-causal accounts.)

[10] Haji and Cuypers say:

> It should be noted that Mele's account of the sufficient conditions for incompatibilist (or libertarian) free action includes a condition that precludes the sort of manipulating that Max does. Mele employs the locution 'compulsion*' to refer to compulsion *not arranged* by an agent. The condition that excludes global and other varieties of covert and nonconstraining manipulation says that the agent (like Jim) has no compelled* nor any coercively produced motivational states.

One advantage of my indeterministic manipulation scenario (that I will introduce in the next section) over Haji and Cuypers's scenario is that while Max's manipulation precludes this condition from being satisfied, Ernie does not suffer compulsion* and so all of Mele's conditions can be satisfied in my scenario.

Clarke's indeterministic scenario (2012: 140) is complicated, but the following summary will suffice for our purposes. Brown is a mad scientist who wants Smith to buy eggs from battery-caged hens rather than the certified organic ones that cost a bit more. Brown watches as Smith chooses which eggs to buy (in a way such that she would satisfy libertarian conditions on free and responsible action when she does choose), and if she chooses the better (certified organic) ones, Brown will use his "Memory Eraser" on Smith such that she must again choose which eggs to buy.[11] Clarke explains:

> In virtue of her libertarian freedom, there is no guarantee that she will make the same choice she did the first time around, just as Brown had no guarantee beforehand that she would make the choice he desired. Now, if Smith persists in choosing the better eggs, Brown will continue wiping her memory and resetting the simulation…Smith is bound to choose the bad eggs eventually. (2012: 140)

When Smith does choose to buy the bad eggs, according to Clarke, we should judge that, although Smith satisfies libertarian conditions on free and responsible action in making her choice, she is not responsible for this choice (since she was manipulated by Brown to make it).

Suppose that Clarke is right that we should judge Smith not responsible for choosing the bad eggs.[12] Even if so, I will now argue that this is not because Smith is manipulated into making

---

[11] Those familiar with van Inwagen's (2000) "Rollback Argument" may find Clarke's argument better associated with the problem of luck for libertarianism than with any manipulation argument, but I will treat it (as Clarke does) as a worry about manipulation. For a discussion of the rollback argument, see Fischer (2012, chapter 6). Interestingly, Pereboom's various discussions of indeterministic manipulation also sound more like luck-related worries than manipulation-related ones. See, for example, Pereboom (2001: 41-54; 2014, chapter 2).

[12] As it happens, I believe that this supposition is false. To see why, suppose Smith chooses the good eggs, has her memory wiped, and then chooses the bad eggs the next time around. I do not find it obvious that she is not responsible for so choosing the second time around. Of course, she is also responsible for choosing the good eggs the first time around, and this might make us less likely to blame her for choosing the bad eggs the second time around, but that's a separate matter. Now

this choice, and thus we should not see Clarke's case as a manipulation challenge to libertarianism. In typical cases of manipulation, the manipulator covertly influenced the manipulated person. But Brown has in no way influenced Smith's values or desires—he has merely removed certain beliefs about her past behavior. Additionally, *contra* Clarke, there is no sense in which "Smith is *bound* to choose the bad eggs eventually" (2012: 140, emphasis added), as the alternative choice remains a robust alternative at each moment of choice, and Brown has no way of shaping or guiding Smith's behavior. Moreover, whereas typical cases of manipulation are those in which the manipulated agent appears not to be the proper *source* of her action, there is no reason to think that Smith is not the source of her choice (given that Brown does not influence her values or desires). So, while Clarke's scenario seems better equipped than Haji and Cuypers's to leave room for the "manipulated" agent to satisfy various libertarian conditions on free and responsible action, Clarke's scenario does not raise a manipulation challenge for libertarians, and certainly not one that is parallel to the manipulation challenge to compatibilism, which is the challenge that this paper is focused on.

Before turning to my new indeterministic manipulation scenario, I want to highlight one important difference between my approach and Clarke's—one that also marks a difference between my argument and Mickelson's, which I mentioned above. In Clarke's case, it may seem to be a matter of luck that Smith chooses as she does, given her circumstances, and thus that she is not responsible.[13] I am willing to grant that there is a challenge to libertarianism implicit in Clarke's case, but, as I see it, this is not a manipulation challenge but rather a luck-based challenge. Mickelson (2015) argues that all manipulation arguments are really only luck-based arguments. I

suppose Smith chooses the bad eggs on the first go-round. Here it is even less plausible that Smith is not responsible, for Brown did absolutely nothing to get Smith to make this choice.

[13] Again, see the works cited in note 11.

disagree with that assessment, though for reasons that are beyond the scope of this paper. Interestingly, however, one could see my project as arguing for the shared conclusion that libertarianism is vulnerable to a parallel manipulation argument but without subsuming worries about manipulation to worries about luck.

4. A New Indeterministic Manipulation Scenario

What would be better, then, in order to provide a challenge to libertarianism that is parallel to the manipulation challenge to compatibilism, is a case of indeterministic manipulation in which it is clear, first, that an agent satisfies the conditions proffered by all types of libertarian accounts and, second, that the agent is manipulated into performing an action on the basis of values or desires that are supplied by another agent. The aim of this section is to construct such a case, and the case will build upon certain earlier cases of indeterministic manipulation that have been introduced to defend compatibilism against the original manipulation argument. The foundation for my manipulation scenario is a case from Stephen Kearns's reply to Mele's Zygote Argument.[14]

Kearns (2012) constructs an indeterministic version of the case from Mele's Zygote Argument. Kearns is concerned with how we are to interpret premise 1 of the Zygote Argument (which says that Ernie is not a free agent and is not morally responsible for anything because of the way his zygote was produced in his deterministic universe) in such a way that premise 2 of the argument can be maintained as well. If we are to understand Ernie's lack of freedom as exclusively explained by the fact that his actions are deterministically caused, then we have lost reference to manipulation and the argument clearly begs the question against the compatibilist. "If one is to show that Ernie's being manipulated is indeed an independent explanation of his lack of freedom,"

[14] I am also building on the case of indeterministic manipulation from Mele's (2005) reply to Pereboom's Four-Case Argument. In Mele's version of the cases, there is a tiny chance that the manipulators will not get the result they want but rather the manipulated agent will be incapacitated.

Kearns explains, "one needs a case in which Ernie is unfree because he is manipulated, but also in which Ernie's actions are *not* deterministically caused" (2012: 384). Kearns then provides the following indeterministic scenario:

> Diana creates a zygote *Z* in Mary. She combines *Z*'s atoms as she does because she wants the zygote to develop into an agent who performs a certain set of actions over the course of his entire life. From her knowledge of the state of the universe just prior to her creating *Z* and the laws of nature of her *indeterministic* universe, she deduces that a zygote with precisely *Z*'s constitution located in Mary will develop into an ideally self-controlled agent, Ernie. As Ernie lives his life, there is a small chance every few seconds that Ernie is incapacitated due to the way Diana created his zygote. If Ernie is never so incapacitated, then he performs that set of actions that Diana has planned. As it happens, Ernie is never incapacitated and performs all those actions Diana has planned. (2012: 385)

Kearns goes on to argue that, even with this modification to the Zygote Argument, the argument fails because, once we have introduced the indeterministic scenario, "we are no longer in a position to insist that Ernie's situation is relevantly similar to a case in which there is no manipulation and in which causal determinism *does* obtain" (2012: 386) and thus premise 2 of the Zygote Argument cannot be maintained alongside the modified premise 1.

Both Mele's reply to the Four-Case Argument and Kearns's reply to the Zygote Argument make use of indeterministic manipulation cases, but these authors aim only to provide an objection to manipulation arguments against compatibilism. Their aim is not to show that libertarian accounts of free will are vulnerable to a parallel manipulation argument. Nevertheless, their indeterministic manipulation scenarios provide a blueprint for a case in which all libertarian conditions on freedom are satisfied by a manipulated agent. Consider the following indeterministic manipulation scenario:

IMS: Diana, a goddess who knows the laws of nature in her *indeterministic* world and who has a very good understanding of human genetics, creates a zygote *Z* in Mary. She combines *Z*'s atoms as she does because she wants the zygote to develop into an agent who performs a certain set of actions over the course of his entire life, culminating in his performing action *A* at time *t30* (when he is 30 years old). From her knowledge of the state of the universe just prior to her creating *Z* (and from her knowledge of the laws of nature), she deduces that a zygote with precisely *Z*'s constitution located in Mary will develop into an ideally self-controlled agent, Ernie. In addition, Diana deduces that, at some time prior to his performing any morally significant decisions or actions, Ernie will have a specific psychological profile, *P*, that includes all of the beliefs, desires, and values that Diana wants for Ernie to possess. From her knowledge of the state of the world and the laws of nature, Diana knows that Ernie's having *P* will make it possible (and more likely than if Ernie has some other psychological profile then) that his life unfolds exactly according to her plan, but if Ernie ever deviates from Diana's plan for his life, she will immediately obliterate Ernie. In addition, Diana frequently but covertly puts Ernie into circumstances that he would otherwise have avoided and that are essential to her plan for Ernie. (Diana's ideas about covertly interfering with Ernie were inspired by the events of the film *The Truman Show*.) To give just one example, on one occasion Diana causes Ernie to become ill just before the start of a party at which Ernie would have met the love of his life (which would have made Ernie's doing *A* at *t30* unthinkable for him). As it happens, Ernie's life goes exactly according to Diana's plan, including Ernie's doing *A* at *t30*, and Ernie is never obliterated.

Furthermore, despite Diana's role in the scenario, Ernie satisfies all compatibilist conditions on free agency (except those conditions which require determinism for free

agency) as well as several (if not all) libertarian conditions on free agency. First and foremost, Ernie is not causally determined by factors beyond his control, since his world is indeterministic and there are, throughout his life, many chances that Diana's plan will fail. At various points in his life, Ernie must choose between acting as is morally required of him and acting in his own perceived best interest, and when he does he influences the character traits that eventually lead to his *A*-ing at *t30*. Just before *t30*, Ernie's intellect represents *A* as the good to be pursued at *t30* (and represents *A* in just the way that Ernie in fact goes about performing *A* at *t30*), and Ernie's volition to *A* at *t30* is formed in consequence of that representation of his intellect. When Ernie does *A* at *t30*, he is doing exactly what he wants when he wants to do it. Whenever Ernie deliberates about what it would be best to do, including during his deliberation about whether to *A* at *t30*, what comes to mind during his deliberation is indeterministically caused to come to mind. Ernie regularly causes action-initiating intentions within himself, thereby bringing about intentions that result in free overt actions, which happens in the case of his doing *A* at *t30*. Ernie's intention to *A* at *t30* has the content *that Ernie performs A in order to satisfy R*, where '*R*' is the reason Ernie does *A*. When Ernie does *A* at *t30*, *A* is caused by both Ernie (the agent) himself and also indeterministically caused by Ernie's having reason *R* for doing *A* at *t30*. Just before *t30*,

Ernie forms a decision to *A* at *t30* such that his making the decision to *A* at *t30* has an actish phenomenal quality, and he is in no way subject to irresistible compulsion.[15]

Since Ernie's circumstances are peculiar, we cannot form a manipulation argument against *libertarianism in general* that would be exactly parallel to the manipulation argument against compatibilism.[16] Still, IMS can be used in an argument schema that will allow for an objection to any particular libertarian account of free will. The structure is as follows, where any particular libertarian account can be substituted for "[libertarian account of free will]":[17]

P1. Intuitively, Ernie is neither free nor responsible for doing *A* (or for his decision to do *A*) at *t30* in IMS.

P2. According to [libertarian account of free will], Ernie is free and responsible for doing *A* at *t30* in IMS.

---

[15] A crucial difference between IMS and so-called "Frankfurt-style cases," named after Frankfurt because of his seminal (1969) presentation of such a case, is that Diana plays a role in the actual sequence which culminates in Ernie's doing *A* at *t30*, whereas the counterfactual intervener in Frankfurt-style cases does not play a role in the actual sequence. (Another importance difference is that IMS aims to leave Ernie with alternative possibilities, whereas the agents in successful Frankfurt-style cases do not possess alternatives.) IMS does, however, incorporate a feature of the so-called "Fischer-type example," which Fischer (2006: 150-151) uses to defend the success of Frankfurt-style cases, in that IMS stipulates that Ernie will be destroyed if he does otherwise than is planned for him by someone else.

[16] King (2013) attempts this (in defense of compatibilism, not to challenge libertarianism), but, as I argue in Cyr (2016), King's parallel manipulation argument fails because, among other reasons, we cannot generalize from an agent's non-responsibility in an indeterministic manipulation scenario to the non-responsibility of all ordinary agents in indeterministic worlds.

[17] Though I have focused on the more popular versions of the manipulation argument, the Four-Case Argument and the Zygote Argument, which aim to show that compatibilism (rather than some compatibilist account of freedom and responsibility) is false, earlier versions of manipulation arguments were aimed at undermining particular compatibilist accounts, especially Frankfurt's (1971) hierarchical account, and these arguments share the same structure as the one I have provided in my own schema. See Slote (1980) for an example of this (though he mainly focuses on coercion, not manipulation), and see Fischer (2006, chapter 12) for a discussion of this kind of attack on his own compatibilist account.

C. Therefore, [libertarian account of free will] is false.[18]

In order to show that this argument schema suffices to raise problems for all extant types of libertarian account of free will, we will need to consider all extant types of libertarian account and to determine whether or not Ernie satisfies those accounts' conditions on free and responsible action. This is the aim of the next section of the paper. Before turning to those accounts, however, it is worth taking stock and asking we should make of this indeterministic manipulation scenario (IMS).

Let me begin by reporting my own reaction to the case. Just as many compatibilists (including McKenna, who takes the hard-line reply to the manipulation argument against compatibilism) do not have the intuition that Ernie lacks freedom and responsibility in the *deterministic* version of the case, I do not have the intuition that Ernie lacks freedom and responsibility in IMS. And that is fine—my aim is not to show that libertarianism is false but rather to show that it is *as vulnerable* to manipulation arguments as is compatibilism. And I do not see a principled reason for treating the deterministic and indeterministic versions of the cases asymmetrically: if one's reaction to the original case of Ernie was to judge that he is not free and responsible, on what basis could one maintain that Ernie is nevertheless free and responsible in IMS?

A natural suggestion is that the indeterminacy in IMS leaves leeway for Ernie to do otherwise than what Diana has planned for him to do, whereas in the deterministic manipulation scenario Ernie is never free to do otherwise than what Diana has planned. Now, it is contentious whether being causally determined to perform some action precludes the (relevant sense of the)

[18] This argument schema parallels an argument schema from Matheson (2018), where Matheson proposes a manipulation argument against compatibilism that omits a generalization premise.

freedom to do otherwise than that action, but let us grant that it does, for the sake of argument.[19] Should one have asymmetric reactions to the types of cases based on whether or not the manipulated agent has the freedom to do otherwise? If the answer is yes, then I take it that the original manipulation argument against compatibilism loses its force—and thus does not reveal a cost of compatibilism—for as long as the compatibilist can address the threat from determinism to the freedom to do otherwise (either by showing their compatibility, as classical compatibilists attempt, or by showing that such freedom is unnecessary for responsibility, as semicompatibilists attempt) there will be no remaining challenge raised by the manipulation scenario. Yet compatibilism is not thought to be uniquely pricey because of the threat from determinism to the freedom to do otherwise; the unique cost of compatibilism is allegedly revealed by the *manipulation* argument. Furthermore, it may be that the feature of cases of deterministic manipulation that typically cultivates a judgment of non-responsibility is the eeriness of being used, unwittingly, by another agent for her own ends. As Neal Tognazzini says, what "haunts" him about compatibilism is the thought that it might imply the compatibility of responsibility and manipulation since, "let's face it, manipulation is creepy" (2014: 358). But notice that, because of the way Diana created Ernie in IMS, and because of her covert interference with him throughout his life, the very same troubling feature of deterministic scenarios is present in IMS as well *even if* Ernie has the freedom to do otherwise in IMS.[20]

[19] Classical compatibilists, such as Vihvelin (2013), would disagree, though see Cohen (2015) and Todd (2017) for recent manipulation arguments that target classical compatibilism in particular.

[20] Another possible suggestion is that Ernie's *sourcehood* is undermined in the deterministic but not the indeterministic version of the case, but insofar as Diana plays the same role in setting up both the deterministic and indeterministic scenarios, it strikes me as implausible to maintain that Ernie is the appropriate source of his actions in only one of the two cases.

IMS appears to be a manipulation scenario in which the manipulated agent nevertheless satisfies extant libertarian conditions on freedom and responsibility. If the case is coherent and libertarianism is, like compatibilism, vulnerable to an argument from manipulation, then manipulation arguments do not reveal a unique cost of compatibilism; instead, if worries about manipulation pose a threat at all, they are a threat to any kind of account according to which we have free will. For my project to be successful (and, in particular, in order to make it clear that libertarianism is vulnerable to a manipulation argument), it is crucial that my case of indeterministic manipulation is indeed a case of manipulation and that it is one in which the manipulated agent does indeed satisfy libertarian conditions on freedom and responsibility. Two potential objections to my argument, then, are 1) that Ernie is not really manipulated to do *A* in IMS and 2) that Ernie does not satisfy libertarian conditions on freedom and responsibility for doing *A* in IMS. Let us take up these objections in reverse order.

One might think that Diana has not left room for Ernie to satisfy all libertarian conditions since she has only left Ernie with the *bare possibility* of the occurrence of events that are contrary to her plan and such bare possibilities are, on some libertarian accounts, insufficient to secure robust possible alternatives for Ernie. In other words, true freedom requires being able to do otherwise in a robust sense, not merely in the sense that it is possible that something else happen to you. But, according to this objection, Ernie's alternatives to doing as he actually does throughout his life are only ever non-robust happenings (in which he is obliterated), so IMS does not allow for Ernie to satisfy all libertarian conditions. But, as I will explain in more detail in the following section, nothing in IMS requires that Ernie only ever has non-robust alternatives. In fact, according to IMS, Ernie *does* possess robust alternatives at various points throughout his life. In order for Ernie to satisfy the conditions on Robert Kane's (1996) account, for example, Diana built various self-

forming actions (SFAs) into Ernie's history, at which times Ernie has robust alternatives even though, had he acted contrary to Diana's plan, he would have been obliterated.

A second potential objection is that IMS is not really a case of manipulation. This sort of objection is complicated by the fact that there is no standard account of manipulation on which participants in this literature are relying. Given the plausibility of the claim that manipulation is not essentially deterministic, my own view is that the correct account of manipulation will entail that IMS is indeed a case of manipulation. Still, one might think that genuine manipulation *guarantees* that the manipulated agents acts in accordance with the manipulator's wishes, but, since Diana does not guarantee that Ernie does *A* at *t30*, her influence on Ernie does not rise to the level of manipulation. But as we learned from Mele, Kearns, and others who have introduced cases of indeterministic manipulation, manipulation is not essentially deterministic. Because manipulation is not essentially deterministic, one can be manipulated even if there was a chance that the manipulation failed.[21] Moreover, there are several positive reasons for thinking that IMS is a case of manipulation. First, Diana has influenced Ernie's values and desires by creating his zygote in just the way she did. Second, Diana has guaranteed that, by a certain time (t30), either Ernie will have performed that action that Diana wanted him to perform or he will have been destroyed.

[21] Tognazzini defends a related point, claiming:

> Nevertheless, I think it's clear that the libertarian will insist that any manipulation scenario that comes together with a strong enough guarantee (even if it's not 100 percent) that the agent will perform as desired is a scenario in which the condition of sourcehood will not be met. The incompatibilist's worry is not simply that the manipulated agent's action is guaranteed, but rather that the agent's actions and values seem to be at the mercy of someone or something other than himself, and this would be true even if the manipulator only made the action 99 percent probable. (2014: 362, n. 8)

I would add that even in cases in which the manipulator made an action less than 99 percent probable (even far less probable) but also would obliterate the agent if things did not go according to plan, there too the worry is that the manipulated agent's actions are at the mercy of someone or something other than himself.

Finally, by creating Ernie with a plan for his life and by standing ready to intervene should Ernie deter from this plan, Diana has shaped Ernie's life, calling into question Ernie's ultimate sourcehood over his doing *A* at *t30* (in just the way that, in the deterministic version of the case, Ernie's sourcehood is called into question by Diana's role in the case).

5. Libertarian Accounts of Free Will

In this section, I consider a handful of token libertarian accounts and show that Ernie (in IMS) satisfies those accounts' conditions on free and responsible action. The three main types of libertarian accounts of free will are event-causal, agent-causal, and non-causal accounts, and I will discuss one token of each.[22]

5.1 Kane's Event-Causal Account

The most popular type of libertarian account of free will is the event-causal type. On these accounts, the indeterminism required for free will and moral responsibility comes in the form of agent-involving events that indeterministically cause other events to occur. The most widely discussed event-causal account is developed by Robert Kane (1996). On Kane's view, an agent is free in performing some action *A* only if she is ultimately responsible for *A*. In order for agents ever to be ultimately responsible for any of their actions, at least some of their actions must be non-derivatively free, that is, "some voluntary actions (including refrainings) of the agents' life histories for which the agents are responsible…must be undetermined. Let us call these undetermined actions 'self-forming actions' (or SFAs) (taking the liberty of assuming that voluntary refrainings can be called actions)" (1996: 74). Kane gives many examples of types of SFAs, but the one that has received the most attention in the literature is that of choosing between doing what morality requires

[22] Although I do not have the space to demonstrate this explicitly, I have formulated IMS with sufficient detail for Ernie to satisfy the conditions of other token libertarian accounts of free will, including Stump (1990), Mele (1995; 2006), and Clarke (2003).

and what one takes to be in one's own best interest. If it is undetermined whether an agent will *A* or will not *A* at time *t1*, and if *A* is what is morally required of the agent but not *A*-ing is what she takes to be in her best interest, then she can be non-derivatively ultimately responsible for *A*-ing or not *A*-ing, whichever she does.

Now, does Ernie (in IMS) satisfy Kane's conditions on free and responsible action? According to the story, "At various points in his life, Ernie must choose between acting as is morally required of him and acting in his own perceived best interest, and when he does he influences his character traits." Since Ernie is able to perform SFAs (and, of course, is not causally determined by factors beyond his control), Ernie satisfies Kane's conditions on free and responsible action and is free and responsible when he does *A* at *t30*.[23]

5.2 O'Connor's Agent-Causal Account

Unlike *event*-causal accounts, *agent*-causal libertarian accounts of free will require that certain events be indeterministically caused to occur not by events but *by the agent herself* (or, as it is sometimes characterized, *by the substance that is the agent*). On one version of the view, developed by Timothy O'Connor (2000), an agent's free action is her causing an action-initiating intention within herself via her active power as an agent-cause. O'Connor explains:

> This direct causing by agents of states of intention goes like this: parallel to event causes, the distinctive capacities of agent causes ('active powers') are grounded in a property or set of properties. So any agent having the relevant internal properties will *have it directly within*

[23] One complication is that Kane has come to require that, in cases of SFAs, the agent must make dual efforts, one effort favoring *A*-ing and the other favoring not *A*-ing, and she must have control over these dual efforts. The control Kane requires for the dual efforts is not, however, the same robust control required for the SFA itself; the former, Kane says, need only be "a *compatibilist kind of control*…akin to what Fischer and Ravizza [(1998)] also call 'guidance control'" (2011: 404, n. 14). It should be clear that Ernie (in IMS) satisfies this condition too.

> *his power to* cause any of a range of states of intention delimited by internal and external circumstances. (2000: 72)

The event brought about in a free action (though the event itself is not a free action) is the agent-involving event *that the agent has a certain action-initiating intention*. Once an agent freely acts in this way, any intended overt action that results from such an instance of free action counts as a free overt action. Since this agent-caused intention cannot itself be caused by a reason (since this would involve event-, not agent-, causation), O'Connor builds into his view that agent-caused intentions possess content of the form *that one perform an action of type A in order to satisfy reason R*, with the result that the agent-caused intention will include a reason-based explanation for the intended action.[24]

Ernie satisfies each of O'Connor's conditions on free action. Ernie is an agent who regularly causes action-initiating intentions within himself via his active power as an agent-cause, thereby bringing about intentions that result in free overt actions, including Ernie's doing *A* at *t30*. Furthermore, as the story goes, Ernie's intention to *A* at *t30* has the content that Ernie performs *A* in order to satisfy *R*, which is exactly what O'Connor's account requires for Ernie's *A*-ing to be on the basis of Ernie's reasons and to count as a free action.

5.3 Ginet's Non-Causal Account

So far we have been concerned with libertarian accounts of free will according to which free actions are those that are caused in certain ways, either by agent-involving events or by the agent herself (or both). The last type of libertarian account we need to consider denies that free actions are caused, hence such accounts are called *non-causal* accounts. Non-causal libertarians are non-causalists about action-explanation, denying that basic actions are to be explained by reference

[24] For more on O'Connor on reasons and causes, see O'Connor (2000, chapter 5).

to causation by any agent-involving events (such as the agent's having a reason for action). Unfortunately, non-causal libertarians typically do not say much about what more is required for *free* action, but one intuitive idea relevant to manipulation seems to be that a basic mental action, explained non-causally, cannot be produced by a manipulator (external to the agent), and so *actions* cannot be brought about by manipulation, much less *free* actions. This might seem to make it difficult to construct a case in which a manipulated agent satisfies non-causal libertarian conditions on free action, but, as I will argue, IMS is just such a case. To see this, let us consider Carl Ginet's (1990) non-causal account.

On Ginet's view, basic acts are uncaused mental acts, such as the making of decisions, which are described as having "what we may call (for lack of a better term) an *actish* phenomenal quality" (1990: 13). Ginet distinguishes his account from agent-causal ones, but he describes this actish phenomenal quality as it seeming like something is directly brought about by the agent herself. When an agent performs a basic action like making a decision, this action is not to be explained by reference to causation by the agent's reasons for action. What more is required for a basic mental action (such as a decision) to count as a *free* action? Ginet says little about this, but Randolph Clarke (2003: 18-19) notes that Ginet requires at least the following for some action *A*, performed by an agent *S*, to count as free:

1. *A* was not causally determined (Ginet 1990, chapter 5)
2. In performing *A*, *S* is not subject to irresistible compulsion (Ginet 1990: 121)
3. *S* is a rational agent (Clarke 2003: 19, n. 9)

As should be clear, Ernie satisfies each of these conditions. Ernie is a rational agent in an indeterministic world, and *A* is not causally determined by any antecedent event. Since Diana does not bring about Ernie's decision by tampering with his brain or otherwise affecting Ernie's normal

way of making decisions, Ernie's making the decision to *A* at *t30* has an actish phenomenal quality, and he is in no way subject to irresistible compulsion. Ernie's decision to *A* at *t30* satisfies all of Ginet's conditions on free action.

One may doubt whether Ernie really satisfies the conditions on free action of a non-causal account like Ginet's since it is built into IMS that Ernie *causes A* at *t30*. But keep in mind that it is consistent with IMS that Ernie's decision to *A* is *uncaused*, and so it is that decision which Ernie does freely, according to an account like Ginet's. And yet Ernie is manipulated into making the decision to *A* despite his satisfying Ginet's conditions on free action.

This completes our survey of a representative sample of libertarian accounts of free will. Since I have argued that each of these accounts entails that Ernie is free (and responsible, though we have been focusing mainly on freedom) either for his decision to do *A* at *t30* or for his actually doing *A* at *t30*, I have been defending P2 of the following schema, which I introduced earlier:

P1. Intuitively, Ernie is neither free nor responsible for doing *A* (or for his decision to do *A*) at *t30* in IMS.

P2. According to [libertarian account of free will], Ernie is free and responsible for doing *A* (or for his decision to do *A*) at *t30* in IMS.

C. Therefore, [libertarian account of free will] is false.

Insofar as one finds it intuitive that Ernie lacks freedom and responsibility, then, one should take this to be a problem for libertarian accounts of free will. But notice that this is just the situation that the compatibilist is in with respect to the manipulation arguments against compatibilist accounts (such as the Four-Case Argument and the Zygote Argument).

6. Conclusion

I have argued that, when it comes to manipulation arguments, compatibilists and libertarians are in the same boat. Indeterministic manipulation scenarios can be created such that an agent is manipulated (and thus intuitively lacks freedom and responsibility) and yet satisfies several (or all) libertarian conditions on free and responsible action. Manipulation arguments divide those who think we have freedom and responsibility from the skeptics, not the compatibilists from the incompatibilists. This paper has left open the possibility that the skeptics are right and that manipulation arguments undermine *all* accounts of the conditions on free and responsible action, but it has also left open the possibility that, while manipulation arguments target compatibilist and libertarian accounts of freedom alike, neither are undermined by the arguments. Indeed, given the extent of the challenge from manipulation, perhaps it would be best to re-evaluate the initially plausible judgment that *certain* manipulated agents (i.e., the ones who have been manipulated in the incredibly sophisticated way that is required in order to satisfy plausible conditions on freedom and responsibility) are not free and responsible for what they do.[25]

Acknowledgements

For helpful discussion and for comments on previous drafts of this paper, I would like to thank the members of my dissertation committee, John Fischer, Dana Nelkin, Michael Nelson, Derk Pereboom, and Eric Schwitzgebel, as well as Dave Beglin, Gabriel De Marco, Matt Flummer, Andrew Law, Meredith McFadden, Jonah Nagashima, and Jeremy Pober. Thanks also to two anonymous reviewers for this journal.

[25] As an anonymous reviewer points out, I have focused on cases of manipulation that build in all of the history needed to satisfy various historical conditions on freedom and responsibility, and one might think that whereas libertarian accounts with historical conditions fare no better than rival compatibilist ones, *nonhistorical* libertarian accounts do fare better than rival compatibilist ones when presented with more invasive (and less historical) cases of manipulation. My own view about these sorts of manipulation cases—such as the Ann/Beth case from Mele (1995: 145; 2006: 164-166)—is that the agent's freedom and responsibility is significantly mitigated, and this is so regardless of whether the case is deterministic or indeterministic. I suppose a committed (nonhistorical) libertarian may make different judgments about the cases, and that would be an interesting line of response to explore, but I leave that to be addressed more fully in future work.

References

Barnes, Eric Christian. (2015) 'Freedom, Creativity, and Manipulation'. *Nous*, 49, 560-588.

Clarke, Randolph. (1995) 'Toward a Credible Agent-Causal Account of Free Will'. In T. O'Connor (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will* (New York: Oxford University Press).

Clarke, Randolph. (1996) 'Agent-Causation and Event Causation in the Production of Free Action'. *Philosophical Topics*, 24, 19-48.

Clarke, Randolph. (2003) *Libertarian Accounts of Free Will*. New York: Oxford University Press.

Clarke, Roger. (2012) 'How to Manipulate an Incompatibilistically Free Agent'. *American Philosophical Quarterly*, 49, 139-149.

Cohen, Y. (2015) 'The Manipulation Argument, At the Very Least, Undermines Classical Compatibilism'. *Philosophia*, 43, 291-307.

Cyr, Taylor. (2016) 'The Parallel Manipulation Argument'. *Ethics*, 126, 1075-1089.

Deery, Oisín, and Nahmias, Eddy. (2017) 'Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood'. *Philosophical Studies*, 174, 1255-1276.

Dennett, Daniel. (1978) *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, Vt.: Bradford.

Fischer, John Martin. (1994) *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell.

Fischer, John Martin. (2006) *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.

Fischer, John Martin. (2011) 'The Zygote Argument Remixed'. *Analysis*, 71, 267-272.

Fischer, John Martin. (2012) *Deep Control: Essays on Free Will and Value*. New York: Oxford University Press.

Fischer, John Martin, and Ravizza, Mark. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, Harry. (1969) 'Alternate Possibilities and Moral Responsibility'. *Journal of Philosophy*, 66, 829–839.

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

Ginet, Carl. (1990) *On Action*. Cambridge: Cambridge University Press.

Haji, Ishtiyaque, and Cuypers, Stefaan. (2001) 'Libertarian Free Will and CNC Manipulation'. *Dialectica*, 55, 221-238.

Kane, Robert. (1996) *The Significance of Free Will*. New York: Oxford University Press.

Kane, Robert. (2011) 'Rethinking Free Will: New Perspectives on an Ancient Problem'. In R. Kane (ed.), *The Oxford Handbook of Free Will*, 2nd ed. (New York: Oxford University Press).

Kearns, Stephen. (2012) 'Aborting the Zygote Argument'. *Philosophical Studies*, 160, 379-389.

King, Matt. (2013) 'The Problem with Manipulation'. *Ethics*, 124, 65-83.

Matheson, Benjamin. (2018) 'The Threat from Manipulation Arguments'. *American Philosophical Quarterly*, 55, 37-50.

McKenna, Michael. (2008) 'A Hard-line Reply to Pereboom's Four-Case Manipulation Argument'. *Philosophy and Phenomenological Research*, 77, 142-159.

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

Mele, Alfred. (2005) 'A Critique of Pereboom's 'Four-Case Argument' for Incompatibilism'. *Analysis*, 65, 75-80.
Mele, Alfred. (2006) *Free Will and Luck*. New York: Oxford University Press.
Mickelson, Kristin. (2015) 'The Zygote Argument Is Invalid: Now What?' *Philosophical Studies*, 172, 2911-2929.
O'Connor, Timothy. (2000) *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
Pereboom, Derk. (2001) *Living Without Free Will*. Cambridge: Cambridge University Press.
Pereboom, Derk. (2014) *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
Slote, Michael. (1980) 'Understanding Free Will'. *Journal of Philosophy*, 77, 136-151.
Stump, Eleonore. (1990) 'Intellect, Will, and the Principle of Alternate Possibilities'. In M. Beaty (ed.), *Christian Theism and the Problems of Philosophy* (Notre Dame: University of Notre Dame Press).
Todd, Patrick. (2011) 'A New Approach to Manipulation Arguments'. *Philosophical Studies*, 152, 127-133.
Todd, Patrick. (2013) 'Defending (a Modified Version of) the Zygote Argument'. *Philosophical Studies*, 164, 189-203.
Todd, Patrick. (2017) 'Manipulation Arguments and the Freedom to do Otherwise'. *Philosophy and Phenomenological Research*, 95, 395-407.
Tognazzini, Neal. (2014) 'The Structure of a Manipulation Argument'. *Ethics* 124: 358-369.
van Inwagen, Peter. (2000) 'Free Will Remains a Mystery'. *Philosophical Perspectives*, 14, 1-19.
Vihvelin, Kadri. (2013) *Causes, Laws, and Free Will*. New York: Oxford University Press.
Waller, Robyn Repko. (2014) 'The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument'. *Philosophia*, 42, 209-222.
Watson, Gary. (1987) 'Free Action and Free Will'. *Mind*, 96, 145-172.
Watson, Gary. (1999) 'Soft Libertarianism and Hard Compatibilism'. *Journal of Ethics*, 3, 351-365.