Review of Analytic Philosophy
Vol. 1 No. 1 (2021)

Research article

An Argument against the Methodology of the Manipulation Argument

Shohei Takasaki

Abstract:
This paper critically examines one of the most influential arguments against compatibilism—"the Manipulation Argument" (henceforth MA), which has been vigorously defended by R. Kane (1996), D. Pereboom (2001, 2014), and A. Mele (2006). MA claims that agents in a deterministic world are, with respect to moral responsibility, relevantly similar to agents whose actions, decisions, or processes of acquiring their character were covertly manipulated by other agents. It will be argued that MA fails to refute compatibilism. The argument, if it succeeds, is important because it can apply to any MA; that is, the argument does not depend on the specific description of manipulation cases or on specific ways of supporting each premise of MA.

Keywords:
Free will, Moral responsibility, Manipulation argument, Source incompatibilism, Methodology

1. Structure of the Manipulation Argument

My aim in this section is (i) to present the structure of the Manipulation Argument (MA) as an argument against compatibilism, (ii) to explain the relation between MA and source incompatibilism by extending MA to an argument for source incompatibilism, and (iii) to state my basic assumption about the methodology of MA.

MA is a template for a kind of argument that aims to show that compatibilism is false by means of "manipulation" cases. The basic structure underlying MA can be stated as follows.

Received May 31, 2020; accepted February 8, 2021; online published March 26, 2021; https://doi.org/10.18494/SAM.RAP.2021.0002

The argument begins by describing a case in which an agent is covertly manipulated in some manner while satisfying all conditions sufficient for the Compatibilist-friendly Agential Structure (CAS).[1] Let us call this case with a manipulated agent S "Case M". Then MA proceeds:

(1) In Case M, agent S is not morally responsible for his action.

(2) S in Case M is no different in any respect relevant to moral responsibility from the agent S under normal deterministic conditions (let us call this case "Case D").

(3) Therefore, the agent S in Case D is not morally responsible for his action.

Because Case D is supposed to be a normal case of action in a deterministic world, we can conclude from (3) that no deterministic agent is morally responsible for his action, which means the denial of compatibilism. Because MA does not depend on the specific account of CAS, MA can be regarded as a general objection to compatibilism.[2]

Here are examples of Case M (Case 2) and Case D (Case 4) from Pereboom (2014):[3]

> Case 2: Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic […] and at times strongly so, with the intended consequence that in his current circumstances, he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first- and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire.[4] (2014, 77, slightly altered)

---

1 I borrow this term from M. McKenna (2008). "CAS is meant by compatibilists to exhaust the freedom relevant condition for moral responsibility" (2008, 142).

2 Responses to MA have taken two main forms: hard-line reply and soft-line reply (this terminology is in McKenna (2008)). Hard-line compatibilists (Frankfurt (2002), McKenna (2008)) reject premise (1), and soft-line compatibilists (Fischer (2004), Baker (2006), Demetriou (2010), Waller (2014), Barnes (2015)) reject premise (2). Because I commit myself neither to the denial of (1) nor the denial of (2), my reply to MA is neither soft-line nor hard-line. I would like to name my position "fundamental-line". In my understanding, Kearns (2012), King (2013), and Schlosser (2015) can also be located in fundamental line. As for a response to King's paper, see Cyr (2016).

3 This numbering is from Pereboom (2014). See also footnote 5.

4 This is one of four cases Pereboom describes; this is why his argument is called the "Four-Case Argument". Some might complain that stating only Case 2 as Case M is unfair to Pereboom because

> Case 4: Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Case 2). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first- and second-order desires. The neural realization of Plum's reasoning process and decision is exactly as it is in Case 2; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill. (2014, 79, slightly altered)

It is worth noting that the plausibility of MA crucially depends on the description of Case M. If, on the one hand, one designed a manipulation so radical as to emphasize non-responsibility of the agent, it might become easier for compatibilists to deny premise (2) (or to argue that the agent in Case M does not have CAS). On the other hand, if one designed a manipulation so trivial as to secure the truth of (2), this would reduce the intuitive plausibility of premise (1).[5] Because of this difficulty of describing an adequate case of manipulation, much ink has been spilled in disputes over whether there is a dialectically appropriate manipulation case, and if there is, what kind of case it is.[6] Fortunately, we do not have to enter this debate, for my argument against MA holds even if we suppose the ideal Case M.

Let us grant (1) for the sake of argument. How do proponents of MA argue for (2)? To see why (2) is supposed to be justified, we must focus on an incompatibilistic intuition that underlies and motivates (2).

Contemporary incompatibilists are divided into two parties in accordance with what aspect of free will or moral responsibility is supposed to be incompatible with determinism: leeway incompatibilists and source incompatibilists.[7] The former think that the ability to do otherwise is necessary for free will and moral responsibility and is incompatible with determinism. The

---

that way of reconstruction would undermine the dialectical power of the original Four-Case Argument. My reply: the way of designing four cases, from a radical manipulation case to a normal deterministic case, would make premise (2) plausible; however, my argument developed in the subsequent section does not rely on the way (2) is supported. In other words, my argument, if it succeeds, applies to the idealized version of MA, which has well-supported premise (2).

5 K. Demetriou (2010) articulates this problem concerning the description of manipulation cases as a form of dilemma, which she calls "the causal control dilemma".

6 Some examples: Fischer (2004), Baker (2006), McKenna (2008), Demetriou (2010), and Matheson (2016).

7 I borrow this terminology from Pereboom and McKenna (2016). This distinction is "rough" in that I don't mean it is mutually exclusive or jointly exhaustive. For instance, Kane (1996) might be regarded as both a leeway and source incompatibilist, because he requires a kind of sourcehood

latter hold that agents cannot be the source of their action in a deterministic world and the lack of sourcehood entails the lack of free will and moral responsibility.

MA is an argument endorsed mainly by source incompatibilists.[8] According to their view, if our action is causally determined by factors beyond our control, we are not the source of the action, which means that the action is not what we do freely. Source incompatibilists would say that an agent in a deterministic world is analogous to an agent who is covertly manipulated by neuroscientists.

Given that MA is motivated by a conception of incompatibilistic sourcehood, it is not surprising that MA as the argument against compatibilism can be extended to an argument for source incompatibilism. One promising way of such extension would be to appeal to inference to the best explanation:[9]

(4) The best explanation of the fact that both the agent in Case M and the agent in Case D are not morally responsible is that, in both cases, the agent's action or decision is deterministically caused by factors beyond his control.

(5) Therefore, if an action or decision is deterministically caused by factors beyond the agent's control, the agent is not morally responsible for his action.

Given this extended version of MA, we can regard the former parts (1)–(3) as the negative argument against compatibilism, and the latter parts (4) and (5) as the positive argument for source incompatibilism.[10] Because my argument attacks the former parts, let us limit our concern to MA as the negative argument.

Before moving on to my critique of MA, I would like to state the main assumption of my argument. I assume that the goal of MA is to convince theoretically neutral but sufficiently rational agnostics, who have no commitment to either compatibilism or incompatibilism. That

condition for free will and moral responsibility, while he holds that that condition requires alternative possibilities at some point.

8 For example, see Pereboom (2001).

9 This is exactly what Pereboom (2014) does: "The salient factor that can plausibly explain why Plum is not responsible in all of the cases is that in each he is causally determined by factors beyond his control to decide as he does. This is therefore a sufficient, and I think also the best, explanation for his non-responsibility in all of the cases" (2014, 79).

10 This distinction would enable us to answer the question of whether a best-explanation argument is necessary for a successful MA, which has been raised in the current literature (cf. Mele (2006), Mickelson (2015), Matheson (2016)). The answer is yes if MA is construed as an argument for source incompatibilism and no if MA is construed as an argument against compatibilism. For a different view from mine, see Mickelson (2015).

is, I assume that if MA succeeds in persuading agnostics to accept (1)–(3), proponents of MA are entitled to claim a victory over compatibilism. I take this assumption to be modest and fair to proponents of MA. This proposal is modest in that it does not require MA to convince compatibilists (which would be very difficult, or, arguably, even an impossible task). Moreover, advocates of MA unanimously agree with this suggestion; indeed, Mele writes: "A more suitable audience for the question about premise (1) […] might be people who have thought long and hard about freedom and moral responsibility and are agnostic about compatibilism" (2006, 190).[11] Hereafter, I will take this assumption for granted.

2. Why the Manipulation Argument Fails

In this section, I will show that MA cannot in principle (i.e., in a way without depending on the detail of the cases) succeed. My argument proceeds as follows. Suppose that an agnostic comes to judge (1) to be true for some set of reasons R (and only R). Then she comes to accept (2) for some reasons suggested by proponents of MA. However, as I shall argue, the agnostic cannot reasonably conclude (3), for the truth of (2) is incompatible with the fact that R is adequate evidence for accepting (1), which means that the agnostic must withdraw her original judgment that (1) is true. I will develop this argument in detail.[12]

Let us begin at the first step of MA. Advocates of MA present the scenario to an agnostic and ask her to judge whether or not S is morally responsible for his action. For what reasons would an agnostic judge that S does not have moral responsibility? It seems that she is expected to mention some conditions particular to manipulation cases (e.g., Case M), that is, a response such as "because the agent's action was caused by the manipulator's intention that the agent act that way" or "because who has responsibility (and is blameworthy) is not S but the manipulator". That is, the conditions mentioned here do not apply to the ordinary deterministic cases (e.g., Case D).

Proponents of MA can, and indeed do, endorse the claim that an agnostic is expected to judge (1) to be true for the set of reasons that mentions the conditions particular to manipulation cases. As Pereboom's remark suggests, it is exactly the way that MA intends:

---

11 Other advocates of MA, for example, Pereboom and Todd, also agree with this assumption.

12 The argument of M. Schlosser (2015) is basically on the same track as mine, though I develop the argument in a different way. (He offers a critique in a form of dilemma that he calls alternative dilemma.) I think my argument is significantly different in two ways. First, I assume that MA is an argument that aims to convince agnostics, which makes, I hope, my critique more general and stronger. Second, I explicitly state and defend an epistemic principle that is needed for showing that MA fails.

> The way manipulation arguments aim to remedy [compatibilists'] putative shortcoming is by first devising a deterministic manipulation case with the hope that it will be more successful at eliciting a non-responsibility intuition than causal determination alone does. (2014, 80)

To put it succinctly, according to Pereboom, Case M is designed to cause an intuition of non-responsibility. P. Todd (2013), one of the proponents of MA, explicates this feature of the dialectic of MA in greater detail. According to him, it is important to distinguish bringing out the judgment that the agent is not morally responsible from making it the case that the agent is not morally responsible. He writes (here, Diana is the manipulator and Ernie is the manipulated agent in the manipulation case from Mele (2006)):[13]

> […] [T]he 'addition' of Diana to a 'normal' scenario involving Ernie can be relevant to bringing out the judgment that Ernie is not responsible. However, this is not to say that the proponent of the argument says that the 'addition' of Diana to such a scenario is itself relevant to Ernie's responsibility. (Todd 2013, 195)

For proponents of MA, the conditions specific to manipulation cases are not relevant to the fact that the agent is not morally responsible. Nevertheless, those conditions should be relevant to the audience's judgment that the agent is not morally responsible, insofar as manipulation cases are supposed to make a non-responsibility intuition vivid. Thus, it seems plausible, and fair to proponents of MA, to suppose that agnostics are expected to judge (1) to be true for the reasons particular to manipulation cases.[14]

Suppose that an agnostic judged that (1) is true for the set of reasons R that mentions the conditions particular to manipulation. The next step of MA is to make the agnostic accept (2). How do proponents of MA try to convince her? Pereboom's strategy in his four-case argument is, again, to appeal to the audience's intuition. It is not clear that this is the best way of

13 Pereboom (2014) agrees with this reading of MA. After quoting Todd's remark, Pereboom says "[h]ere is the dialectic as I see it, which accords with Todd's assessment".

14 Some qualifications are in order. First, I assume that the relation between a judgment and a justifying reason is causal, though I take it that my argument does not rest on the particular theory of an epistemic-basing relation. For instance, my argument would hold given Counterfactual Theories or Doxastic Theories of the basing relation (cf. Swain 1981, Tolliver 1982). Second, it should be noted that a cause of one's belief might not always be a reason to believe it, such as in cases of a deviant cause (I thank an anonymous referee for making this point clear). Nevertheless, I assume, in the present discussion, that in the acceptance of premise (1), a cause of an agnostic's belief that (1) is true, i.e., the manipulation, is also a reason for holding that belief. For, as I argued, she would in fact cite the manipulation as the reason for holding her belief that (1) is true.

supporting (2). It might be possible to provide more "dialectical" support for (2), and thereby succeed in persuading an agnostic. In any case, my argument does not depend on the specific way of convincing her to accept (2). For the sake of argument then, let us suppose that she comes to judge that (2) is true for some reason.

Here is the gist of my critique. Recall that (2) is the thesis that S in Case M is no different in any respect relevant to moral responsibility from the agent S who acts in Case D. This entails that R, which mentions the conditions particular to manipulation, is irrelevant to the agent's moral responsibility. Therefore, (2) is not compatible with the claim that R is a good reason for judging that (1) is true. It follows that the agnostic, qua a rational subject, must withdraw her original judgment to accept (1), given that she has no other reasons to accept (1). However, this means that she has no reason to accept (3), for she cannot conclude (3) unless she sustains both (1) and (2).

My argument rests on the following epistemic principle, which is supposed to hold between a judgment and a set of reasons supporting that judgment:

(E) If (i) the subject S judges that P for the set of reasons R at t_0, (ii) S judges that R is irrelevant to the truth of P at t_1 ($t_0 < t_1$), and (iii) S has no other reasons for the judgment that P at t_1, then S should withdraw the original judgment that P after t_1.

This principle is highly plausible. For instance, suppose that John eavesdropped on his fellow workers' conversation and judged that his supervisor's partner is vegetarian. However, afterwards, John learns that the person they talked about was not his supervisor but another professor. Then, if he had no other reasons for his judgment that his supervisor's partner is vegetarian (perhaps the professor never talks about his private life), then John must withdraw his original judgment. If he still held that belief, we would regard him as irrational.

An agnostic seems to be in a similar epistemic situation as John. An agnostic judges (1) to be true; that is, she judges that S in Case M has no moral responsibility on the basis of some conditions (= R) particular to manipulation. After that, she accepts (2), which says that S in Case M is no different in any respect relevant to moral responsibility from the agent S in Case D. This entails that conditions particular to manipulation are irrelevant to the agent's moral responsibility. Then she, qua a rational agent, should become aware that R is irrelevant to the truth of (1). Moreover, since she is ex hypothesi theoretically neutral on the debate, she does not have any reasons other than R for judging (1), even after she accepts (2). Therefore, by principle (E), she must withdraw her original judgment that (1) is true. This suggests that she should remain agnostic about (3).

Proponents of MA might reply that condition (iii) in (E) is not satisfied, because in accepting (1) and (2), an agnostic is expected to realize that the better reason why S in Case M is not morally responsible is causal determination (or the lack of sourcehood it implies). However, this response is not available to proponents of MA, for we can obtain this reason only after premise (4), the premise of the argument for source incompatibilism, is established. In other words, to make an agnostic accept (1) and (2), advocates of MA cannot appeal to "incompatibilist sourcehood", for otherwise they would beg the question against compatibilism. Thus, an agnostic has no reasons for accepting (1) other than R.

3. Objections and Replies

I will now address three possible objections to my argument. First, one might claim that an agnostic's intuition about Case M is a "bare intuition", which is, in principle, inexplicable.[15] In other words, she might just accept (1) for no reason. I think this claim is implausible. I concede that sometimes bare intuition might provide a strong reason to accept or reject a particular philosophical thesis,[16] but I don't think that an intuition involving Case M is a bare intuition. First, as a matter of fact, if the agnostic were asked the reason why she judged (1) to be true, she would reply by mentioning some conditions particular to manipulation. Moreover, that kind of agnostic's response is exactly what the proponents of MA intend to invoke, for, as I argued before, they design manipulation cases in the hope that conditions particular to manipulation cause the agnostic's judgment of non-responsibility.

Note that I don't claim that intuitions involving Case M, whether they are "bare" or not, cannot have any evidential force. Rather, so far as the agent in Case M satisfies all compatibilist-friendly conditions, they should be regarded as strong evidence against the compatibilists' proposal. My point is that intuitions involving Case M, when combined with the acceptance of (2), does not support conclusion (3).

Second, one might reconstruct MA in a different way than I construed.[17] According to this reconstruction, when we accept (1), the manipulation is wrongly taken to be the cause of the judgment, and once we accept (2), we come to see that the real ground of the judgment was not

15 Fischer (2016) also considers this issue.

16 Intuitions involving Gettier cases in epistemology and a Gödel-Schmitt case in philosophy of language might be successful instances of bare intuitions. Thanks to an anonymous referee for making me recognize this point.

17 Thanks to an anonymous referee for pressing this point.

the manipulation per se, but the causal determination in general. The latter is a more encompassing explanation of why the agent is not responsible for her action.

Surely, we often wrongly recognize the cause of our own judgment and revise our belief about the cause without changing the judgment itself. This kind of revision in belief, however, is justifiable only if we obtain another, independent reason supporting the original judgment. Otherwise, we should withdraw our original judgment, as principle (E) tells us. If my argument in the previous section is correct, what the acceptance of (2) entails is just that the ground of the judgment about (1) is not a good one. The acceptance of (2) itself does not provide any further ground for the claim that the agent is not morally responsible because of the causal determination. Rather, this claim should be established through arguing for (4). If, by accepting premise (2), one comes to be convinced that causal determinism itself undermines moral responsibility, then incompatibilists would not need premise (1) in the first place!

Third, we should consider the possibility of the agnostic's "conversion" to incompatibilism. Perhaps an agnostic, who originally was "theoretically neutral", switched her view to incompatibilism in the process of accepting (2). That is, once she accepts (2), she might realize that her original reason R is an incorrect one and instead hold (1) for another reason (say, causal determination). Why can't the idealized version of MA have such a power of conversion?

This objection overlooks the fact that (2) does not provide any reason to accept incompatibilism. What (2) claims is only that the agents in Case M and Case D are no different with respect to moral responsibility. Taken on its own, (2) leaves open the possibility that both agents do have the same amount of moral responsibility.[18] Therefore, (2) has no power to convert an agnostic's opinion to incompatibilism.[19]

If my argument is sound, we can conclude that MA fails to refute compatibilism in principle. Even if we suppose an idealized MA, that is, the one that has a well-described Case M and premise (2) with compelling intuitive support, my argument can apply to it.

4. Concluding Remarks

If my argument is sound, MA cannot be regarded as a successful refutation of compatibilism. That said, I don't think that MA has no significant insight against compatibilism. Premise (1) as an independent claim can be an important challenge to compatibilism, insofar as the agent S seems to have CAS but seems not to be morally responsible for his action. As far as

18 That is, there is a possibility of the "hard-line" reply here: the agnostic may instead conclude that the agent in Case M and the agent in Case D are both morally responsible.

19 K. Vihvelin (2017) mentions a similar objection.

there is such a case, compatibilists are required to revise their own accounts of moral responsibility, or to admit the consequence that the agent in that manipulation case is morally responsible.[20] In this sense, compatibilists can make use of manipulation cases to improve their theory.

My argument also has a significant consequence for source incompatibilism. Recall the extended version of MA (i.e., (1)–(5)). In this argument, the thesis of source incompatibilism is drawn by "the inference to the best explanation", given that both agents in Case M and Case D are not morally responsible. However, if my critique is right, source incompatibilists cannot establish their positive claim in this way.

In my estimation, the reasoning of MA as a positive argument for source incompatibilism is the other way around. One of the flaws of MA I have shown is that, in order for proponents of MA to justify (1) and (2), they cannot make use of what they think is the ground of (1) and (2) (i.e., causal determination) without begging the question against compatibilism. The root of the problem seems that they put their theoretical ground of (1) and (2) on (4). Therefore, I think (1) and (2) should be regarded not as the premises of the argument for source incompatibilism, but as the consequences of source incompatibilism. What is required for source incompatibilism is an a priori argument that purports to show that "sourcehood" is necessary for free will and moral responsibility and it is incompatible with determinism.

References

Baker, L. R. (2006). Moral Responsibility without Libertarianism, *Noûs* 40, 307–330.

Barnes, E. C. (2015). Freedom, Creativity, and Manipulation, *Noûs* 49, 560–588.

Cyr, T. (2016). The Parallel Manipulation Argument, *Ethics* 126, 1075–1089.

Demetriou, K. (2010). The Soft-Line Solution to Pereboom's Four-Case Argument, *Australasian Journal of Philosophy* 88, 595–617.

Fischer, J. M. (2004). Responsibility and Manipulation, *Journal of Ethics* 8, 145–177.

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

Frankfurt, H. (2002). Reply to J. M. Fischer, S. Buss, and L. Overton eds., *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, MA: MIT Press.

Haji, I. and Cuypers, S. (2006). Hard- and Soft-Line Responses to Pereboom's Four-Case Manipulation Argument, *Acta Analytica* 21, 19–35.

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

Kearns, S. (2012). Aborting the Zygote Argument, *Philosophical Studies* 160, 379–389.

20 In other words, compatibilists can take either a soft-line or hard-line reply for each manipulation case, depending on the details of the case (for a similar suggestion, see Demetriou 2010).

King, M. (2013). The Problem with Manipulation, *Ethics* 124, 65–83.

Matheson, B. (2016). In Defence of the Four-Case Argument, *Philosophical Studies* 173, 1963–1982.

McKenna, M. (2008). A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument, *Philosophy and Phenomenological Research* 77, 142–178.

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

Mickelson, K. (2015). The Zygote Argument Is Invalid: Now What?, *Philosophical Studies* 172, 2911–2929.

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

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

Pereboom, D. and McKenna, M. (2016). *Free Will: A Contemporary Introduction*, NY: Routledge.

Schlosser, M. (2015). Manipulation and the Zygote Argument: Another Reply, *Journal of Ethics* 19, 73–84.

Swain, M. (1981). *Reasons and Knowledge*, Ithaca, NY: Cornell University Press.

Todd, P. (2013). Defending (a Modified Version of the) Zygote Argument, *Philosophical Studies* 164, 189–203.

Tolliver, J. (1982). Basing Beliefs on Reasons, *Grazer Philosophische Studien* 15, 149–161.

Vihvelin, K. (2017). Arguments for Incompatibilism, Stanford Encyclopedia of Philosophy (from the Fall 2017 edition), E. N. Zalta ed. https://plato.stanford.edu/entries/incompatibilism-arguments/ (accessed March 2018)

Waller, B. (2014). *The Stubborn System of Moral Responsibility*, Cambridge, MA: MIT Press.

About the Author

Shohei Takasaki received his B.S. degree from The University of Tokyo, Japan, in 2013 and his M.S. degree from The University of Tokyo in 2015. Since 2020, he has been a lecturer at IUHW Shioya Nursing College. His research interests are free will and moral responsibility.

✉ shoheitakasaki72@gmail.com