

I am very excited to have the opportunity to comment on Haji's piece, as it is rich with lessons for those interested in manipulation arguments. In broad strokes, Haji tackles three of the most central issues currently surrounding such arguments:

- (A) The proper interpretation of Manipulation Argument (i.e. the formal structure shared by all manipulation arguments), and the relation between it and particular manipulation arguments.
- (B) Michael McKenna's recent distinction between 'hard-line' and 'soft-line' response strategies, and the soundness of his argument that every 'credible' manipulation argument (including Pereboom's Four-Case Argument) requires a hard-line reply.
- (C) The dialectical role of "ultimate origination" principles in the success of manipulation arguments, especially that of Pereboom's *Principle O* in the Four-Case Argument.

Haji's chapter is detailed and complicated, so there are many interesting points which I could not address in my comments. I look forward to having them come up in our discussion. My goal here is to stir things up, so in the following I suggest a few concerns one might have with some of Haji's views on each of the key issues above. For simplicity, I will not explicitly discuss event-causal libertarianism, though I believe that the comments below will be equally relevant to proponents of such a view.

(Hereafter, I will number each paragraph for ease of reference.)

Issues concerning Pereboom's Principle *O*

[4] "It would be dialectically unfruitful" to understand Principle *O* as playing the role of a premise in the 4-CA, Haji says, for "if a compatibilist or libertarian doesn't buy *O*, then...she won't regard the four-case argument, so interpreted, as compelling" (Haji, 28). This is true, of course, but a proponent of the Four-Case Argument (hereafter, '4-CA') would likely point out that it would also be a serious misrepresentation to posit *O* as a premise in this manipulation argument.

[5] Pereboom does not invoke the truth of *O* in order to generate the power of his argument, the proponent might argue, but offers it as a final *diagnosis* of why his argument is so successful. Pereboom's preliminary diagnosis of the common intuitive responses to (at least) Case 1 and Case 2 is that manipulation undermines Plum's freedom and moral responsibility because Plum is causally determined to think and behave as he does by "alien-deterministic events" (Pereboom, *Living Without Free Will*, 116). Indeed, it is only in virtue of this diagnosis that Pereboom is able to reach the extremely strong conclusion that *no* compatibilist account could ever succeed—his diagnosis indicates that the responsibility-undermining feature of manipulation cases is determinism itself, which is inconsistent with the truth of compatibilism. Pereboom then goes a step farther, concluding the 4-CA with an attempt to explain why causal determination should have this freedom- and responsibility-undermining effect. Here he finally suggests that the best explanation is the truth of Principle *O* (Pereboom, 126).

[6] If this view of the role of *O* is correct, then Haji would be right to say that the 4-CA is not a *deductive* argument in favor of *O*, but it would be wrong to overlook that Pereboom is offering an *abductive* argument in its favor. Thus, without a response to Pereboom's argument to the best explanation, it is not open for a compatibilist to 'not buy *O*'; in light of the 4-CA, the compatibilist can reject *O* only when they have an alternative diagnosis of equal or greater plausibility *in hand*. So, while it is certain that Pereboom's belief in the truth of a principle like *O* played a guiding role in his construction of the 4-CA, his argument does not *presuppose* the truth of *O* in a problematic way. Quickly, then, consider Haji's three worries with Principle *O*:

[7] *O* conflicts with some of his intuitive responses to other cases (Rosa/Roselle & God).

REPLY: While I think that intuitive reactions which are inconsistent with the truth of *O* might give the compatibilist hope that there is an alternative to Pereboom's diagnosis out there, the compatibilist is not justified in rejecting *O* until she actually shows the error in Pereboom's abductive argument. (See Haji, 29-30)

[8] *O* cannot discriminate between intuitively distinct cases (indoctrination vs. normal learning).

REPLY: If *O* is true—which we must assume until a successful reply to the 4-CA is provided—there is no morally relevant difference between a normal upbringing and an indoctrination-laden upbringing in either deterministic or indeterministic worlds (assuming the individuals in question are not agent-causes). Once again, I think that a person might find that the intuitive difference here suggests that a reply to the 4-CA will be found, but the intuitive appearance of a principled difference on its own does not provide any justification for rejecting *O*. (See Haji, 30-31)

[9] *O* invokes a notion of control that begs the question against the compatibilist.

REPLY: Pereboom claims that, in light of the 4-CA, "we have good reason to believe that an agent cannot be responsible for decisions that are produced by a deterministic process that traces back to causal factors beyond her control—decisions that are alien-deterministic events" (Pereboom, 126). In other words, Pereboom's generalization strategy is designed to illuminate that the relevant account of control, whatever it is, is a type that is incompatible with determinism. Contrary to what Haji claims, then, the anti-compatibilist interpretation of *O* which Pereboom supports does not beg the question against the compatibilist (See Haji, 34-36).

[10] If these responses are right, then Haji's suggestions regarding a compatibilist-friendly principle of origination may be of great use once the 4-CA has been defeated, but none of these worries constitutes an *argument* against the truth of *O*.

Issues Concerning the Form of the Manipulation Argument (MA)

[11] Like McKenna before him, Haji presents MA as a general form or template of an argument for incompatibilism, of which Pereboom's 4-CA is merely an instance (Haji, 1-3). That is, MA is supposed to capture the structural similarities that all manipulation arguments, like the 4-CA and Mele's Ann/Beth argument, have in common. Often, though, Haji represents the relation between MA and the 4-CA in other, seemingly problematic ways:

[12] Haji states that Premise 2 of MA is “predicated on” the 4-CA (Haji, 3), and elsewhere speaks of the 4-CA as a “component” of MA (Haji, 28), as though the 4-CA were an *independent argument* for a particular premise of MA rather than an *instance* of this argument type. While I agree that the 4-CA is unique among manipulation arguments in the degree to which it provides support for the conclusion that there is no principled difference between manipulation victims and the normal determined agent, this is just to say that Pereboom’s argument is stronger than other instances of MA because the 4-CA offers more support for the truth of *its own premises* than do other manipulation arguments.

[13] Perhaps the discussion in paragraph 12 helps to explain why, at other times, Haji speaks of MA and the 4-CA as though they were equivalent. For instance, Haji discusses two different interpretations of the 4-CA (Haji, 25), and then a section later speaks as though he had just presented two interpretations of MA (Haji, 27). For another example, consider Haji’s version of MA. His Premise 2 reads: “Regarding free action and moral responsibility, there is no relevant difference between manipulated S’s A-ing and any action deemed to be free and for which its agent is morally responsible on *any compatibilist account* of free action and moral responsibility” (Haji, 3; italics added). It is true that the 4-CA includes a defense of this premise, but it is not the case that *every* manipulation argument has the lofty goal of undercutting *all* possible compatibilist accounts—Mele’s Ann/Beth argument, for instance, does not. In fact, I am unaware of *any* other manipulation argument in the literature which attempts to demonstrate that compatibilism is, in principle, untenable. Even if there are a handful of others, though, the conclusion Haji attributes to MA that “no compatibilist account of free action and moral responsibility is true” is characteristic of the 4-CA but not of manipulation arguments generally, which means that Haji’s formulation of MA is overly narrow.

[14] Given that Haji’s vision of MA is so tightly bound to the 4-CA, it is not surprising that he posits that Premise 1 of MA “rests on” Pereboom’s Principle *O*, just as he claimed that the 4-CA presupposed the truth of *O* (Haji, 3). However, as discussed in paragraphs 4-10, Haji seemed mistaken in his assessment of the dialectical role played by *O* in the 4-CA, and Haji’s portrayal of the role that Principle *O* plays in MA is even more troublesome. For example, Haji states that “it is not clear whether [Principle *O*] will do the work that it is meant to in the Manipulation Argument”, where one would have expected him to limit his comments on its role in the 4-CA. That is, it appears that there should be no expectation of there being general work for *O* to do in MA, for it is not an essential aspect of a manipulation argument that it invoke or presuppose an ultimate origination principle of any kind—not all manipulation arguments defend *source* incompatibilism.

[15] In light of paragraphs 12-14, one might worry that Haji’s interpretation of MA draws too heavily on the 4-CA. Consider, by contrast, the account of MA offered by McKenna (“A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument”, sec. 1):

1. *If S is manipulated in manner X to A, then S does not A of her own free will and is therefore not morally responsible for A’ing.*

2. *An agent manipulated in manner X to A is no different in any relevant respect from any normally functioning agent determined to do A from [the Compatibilist-friendly Agential Structure, or] CAS.*
3. *Therefore, if S is a normally functioning agent determined to A from CAS, she does not A of her own free will and therefore is not morally responsible for A'ing.*

[16] McKenna's account accommodates a greater variety of manipulation arguments, leaving room for arguments with weaker conclusions than the 4-CA. I will discuss the importance of these issues more below, but would like to move forward under the assumption that the 4-CA is properly seen as a particular *version* of McKenna's MA and focus on the advice Haji provides for responding to it as such.

Issues Concerning Respect for Soft-Line Replies to the 4-CA

[17] Haji, I think, is exactly right when he points out that many compatibilists have already conceded that Pereboom's manipulation victims are not free or responsible, and for *these* compatibilists a hard-line reply is no longer an option (Haji, 26-7). Given the strength of their intuitive responses to Pereboom's cases, they cannot deny Premise 1 of MA in good faith. Therefore, these compatibilists must either identify a principled difference which explains this judgment in a compatibilist-friendly way or admit defeat.

[18] It is also worth emphasizing Haji's sensitivity to the fact that only the subset of compatibilists who have the intuition that manipulation victims are not responsible must confront the central challenge of the 4-CA: finding the interpretation(s) of manipulation which generates this intuition and then providing a soft-line reply (Haji, 27). By contrast, the 4-CA puts no direct pressure on other compatibilists, like Frankfurt and McKenna, who lack the intuition that manipulation is menacing (although I will suggest below that a manipulation argument can put indirect pressure on a compatibilist depending on whether he endorses a hard-line or soft-line reply) (Haji, 24).

[19] One might add that McKenna's suggested strategy, by comparison, seems to leave these soft-line-seeking compatibilists in a very tough spot (see Haji, 18-19). McKenna suggests that compatibilists should seek out interpretations of manipulation on which their account of CAS is satisfied, but he does not address how those who were struck with the *prima facie* intuition that manipulation victims are not morally responsible should go about explaining this response. Following McKenna's strategy, then, the compatibilist might go through interpretation after interpretation of the manipulation on which CAS is clearly satisfied and never have the troubling intuition in response to these interpretations. But, as Haji (rightly) says, only when the compatibilist uncovers the account of manipulation which generated the troubling intuition is she in a position to answer Pereboom's challenge, for only when this account is revealed can she hope to find a principled difference between such manipulation and a normal agent in a deterministic world (Haji, 27).

[20] It may appear, then, that Haji and McKenna are in direct disagreement about how to best deal with manipulation arguments and that Haji emerges victorious. A closer examination of

McKenna's argument against soft-line replies, though, might reveal that Haji's victory, and the nature of the disagreement itself, are less clear-cut than they first appear.

Issues Concerning Respect for Hard-line Replies to MA

[21] The central tenet in Haji's argument against McKenna's rejection of soft-line replies is that McKenna allegedly presupposes that all compatibilist-friendly causal routes are 'deviant' (Haji, 11). Haji provides a plausible argument that this assumption implies that *all* causal routes are deviant, which Haji sees as incompatible with one of the premises in McKenna's argument against the soft-liner, i.e., the assumption that at least *some* causal routes are normal, and thus McKenna's argument fails (Haji, 14).

[22] Since the success of Haji's arguments rests upon his claim that McKenna is committed to the assumption that all compatibilist-friendly causal routes are deviant, I would like to focus on his reasons for attributing this assumption to McKenna. It appears that Haji is worried that the way in which McKenna talks about "credible" manipulation cases is too closely related to what is required of a *successful* manipulation case. McKenna is clearly demanding that any credible case has to succeed in satisfying CAS, for he dismisses those that do not as non-starters (see Haji, 18), and Haji contends that McKenna also requires that a credible case generates the judgment that manipulation is in fact menacing, or equally bad, that the causal route is deviant (Haji, esp. 11 & 17).

[23] If McKenna were truly committed to this understanding of a credible case, surely Haji would be right that McKenna unwittingly undercuts the viability of his own reply to the 4-CA with his attack on soft-liners—and without begging the question against hard-liners (Haji, 17). I am sympathetic to Haji's worry, but this shocking conclusion does invite the question of whether there is a more charitable interpretation of McKenna's argument against soft-line replies than the one Haji suggests.

[24] To draw out a possible alternative, recall paragraphs 11-16 which raised concerns that Haji's version of MA was too heavily informed by features of the 4-CA. One might wonder if a similar type of problem has led to McKenna's conclusion that the 4-CA needs a hard-line reply. That is, perhaps McKenna's attention to the similarities between all manipulation arguments and his many insightful comments about MA can help to explain why he is convinced that any successful response to the 4-CA must end with a hard-line reply.

[25] Backing off from the 4-CA and Haji's framing of McKenna's argument, consider McKenna's claim that "a soft-line reply to a well-crafted version of MA can only temporarily forestall the inevitable" (McKenna, Sec. 2). (Presumably, if MA does not have an invalid form, there can only be replies to its best versions, and not to MA itself; still it is useful shorthand to speak this way.) McKenna is surely right about this, but perhaps he goes astray in his response to the 4-CA because he has started with a focus on what is required to reply to manipulation arguments as a class and then applied the lessons of that inquiry too directly to individual manipulation arguments. That is, it appears that McKenna reasons from the fact that every compatibilist will have to take a hard-line in responding to some version of the MA (because any CAS could be artificially fabricated by some version of manipulation), that it must also be the

case that all compatibilists need to take a hard-line reply to the 4-CA in particular. Unfortunately for McKenna, there seems to be little reason to think that this is a sound inference.

[26] Perhaps McKenna is assuming that the type of manipulation described in the 4-CA's manipulation cases is open to unconstrained interpretation, such that for any given account of CAS, the type of manipulation outlined in the 4-CA's cases could satisfy it. If so, then McKenna's conviction that a soft-line reply to the 4-CA is unavailable would make sense, for any attempt to give a soft-line reply would be followed with a new interpretation of Pereboom's manipulation cases until every failure was remedied; a hard-line would then be required to finish the defense. Looking at Pereboom's description of the manipulation processes, though, it seems that Pereboom means to be describing a very specific type of manipulation, even though it does not take much effort to see that the manipulation is problematically underdescribed. Thus, one will not be able to insert any old account of manipulation into Pereboom's cases and still preserve the type of manipulation that Pereboom describes—and presumably when the type of manipulation he describes is no longer preserved, neither is his 4-CA.

[27] Return now to the makings of a credible case. If McKenna is working under the (mistaken) assumption that an interpretation of any given manipulation case could be developed that would be able to satisfy one's preferred account of CAS without changing the manipulation argument at issue, this would explain why he thinks that the need for a hard-line reply to the 4-CA will remain after all soft-line replies have been exhausted. Under this interpretation of McKenna, it also makes sense that he would tie the notion of credibility to the actual satisfaction of CAS but not to the intuition that the manipulation in question is menacing: since CAS satisfaction is inevitable, only the compatibilist who is able to dismiss her intuition that the manipulation is menacing when CAS is satisfied will be able to defend her theory. Thus, while the credibility of a case might require the *prima facie* intuition that the manipulation is menacing, McKenna's broad view of the 4-CA no longer seems to require, *pace* Haji, that one agree with the proponent of the argument that the manipulation is in fact menacing (Haji, 11).

[28] Perhaps I would need to say more to make this objection to Haji's critique clear, but my suggestion is that Haji attributes an assumption to McKenna that would be plausible on Haji's view of the dialectic but not on McKenna's. I have tried to provide an alternative interpretation that is in keeping with McKenna's overall approach and that does not commit him to the view of credibility Haji attributes to him and, thereby, does not commit him to the assumption that threatens to undermine his own preferred response to the 4-CA. Notably, though, I don't think that McKenna's attack on soft-line replies to the 4-CA looks any more compelling under my interpretation than Haji's.

A Suggestion for a New MA-Response Strategy

[29] Again, as discussed in paragraphs 11-16, Haji seems to understand MA in terms of the 4-CA. As a result, his commentary often suggests that the soft-line response strategy he defends with respect to the 4-CA should be part of the general response strategy to all manipulation arguments. Those, like Frankfurt and McKenna, who are more prone to adopt hard-line replies than many other compatibilists, however, would likely complain that Haji undervalues the place of hard-line replies—and, I believe, this would be a fair complaint.

[30] On the other hand, we have seen that McKenna's suggested strategy appears to be too dismissive of the value soft-line replies to individual arguments. I have suggested that McKenna adopts this line because he overlooks the importance of the distinctions among manipulation arguments. Only by recognizing distinct arguments can we respect Haji's insight that every unique manipulation argument puts unique demands on different groups of compatibilists, depending on their intuitive responses to the manipulation cases and their preferred account of CAS.

[31] It is unfortunate that neither McKenna nor Haji want to emphasize that *both* hard-line and soft-line replies are essential elements to responding to manipulation arguments as a class, and that they try to ignore the fact that both types of reply play a role in their respective response strategies. For instance, when McKenna dismisses all non-starters in the first step of his suggested strategy, he describes these non-starters as those that fail the CAS and thus warrant a rejection of Premise 1 of MA—which, of course, satisfies his definition of a soft-line reply (see Haji, 18). Haji, too, recognizes that compatibilists can give an honest hard-line reply if they lack the intuition that manipulation is menacing, but because he is defending the value of soft-line replies his comments obscure the fact that these hard-line replies can be equally powerful in bringing down a given manipulation argument (see Haji, 24).

[32] On the other hand, Haji and McKenna each offer important insight on what the general strategy for responding to MA must be like. From Haji we have learned that compatibilists must be allowed to pursue a soft-line to any particular manipulation argument if this is what is required to save their view in light of their judgment of the manipulation case. From McKenna, we know that a compatibilist cannot possibly give a soft-line reply to every manipulation argument, so room must also be made for the compatibilist to give a hard-line reply. With a bit of supplementation, then, we might consider adopting a new strategy that gives proper respect to both hard-line and soft-line replies. For instance, consider what I shall call the "Switch-Line Strategy" presented below.

-When a new manipulation argument is given, the compatibilist should proceed as follows:
(Premises 1-3 refer to McKenna's articulation of MA)

Step One: Give your *prima facie* intuition about the freedom and/or responsibility of the manipulated character:

-If your intuition is that the character is free and responsible, then Premise 1 of MA is false; this is a (level-one) **hard-line reply**.

-If your intuition is that the manipulation undermines the character's freedom and/or moral responsibility, proceed to Step 2.

Step Two: Investigate the manipulation case for a failure of your preferred CAS:

-If investigation reveals multiple interpretations of the manipulation are available, specify each and return to Step One.

-If a failure is discovered, Premise 2 of MA is false; this is a (level-one) **soft-line reply**.

-If a failure is not discovered, proceed to Step 3.

Step Three: Review your preferred CAS for a missing condition, one which would reveal a principled difference between the cases:

- If new condition is found, Premise 2 of MA is false; this is a (level-two) **soft-line reply**.
- If new condition cannot be found (one that is not *ad hoc*), proceed to Step 4.

Step Four: Refresh your mind and then run a McKenna-style inverse manipulation argument (McKenna, sec. 7), ending in the manipulation case at issue:

- If you find that your *prima facie* intuition dissolves as you focus on the satisfaction of your account of CAS—such that the manipulation no longer seems menacing—then Premise 1 of MA is false; this is a reflective (level-two) **hard-line reply**.
- If you find that your *prima facie* intuition remains unscathed, proceed to Step 5.

Step Five: Admit defeat; convert to incompatibilism.

[33] Were people to adopt this new strategy, there would no longer be “hard-liners” and “soft-liners”, but only compatibilists who take a hard-line here and a soft-line there. It also leaves open the possibility that the same compatibilist might give a combination of soft- and hard-line replies to the same manipulation argument if there are multiple interpretations of the manipulation available. I believe that this could help to resolve some of the apparent disagreement among compatibilists about how to respond to manipulation arguments, as it recognizes both a common overall strategy as well as the fact that particular responses will vary. I believe it would also have many significant benefits, such as:

[34] Since the Switch-Line Strategy offers two levels for each type of reply (soft and hard), it allows one to follow McKenna’s advice to start by dismissing all manipulation arguments that are obviously not in the running for satisfying one’s account of CAS, while leaving open the possibility that a closer look at the manipulation will reveal that CAS is not satisfied, as Haji would like. Thus, it respects Haji’s insight that sometimes the last step in resolving a manipulation argument will have to be a soft-line reply, while acknowledging that McKenna is right to think that that sometimes the final reply must of a hard-line variety.

[35] It allows observers to track the variety of responses given to each manipulation argument by different compatibilist contenders. While valuable to committed theorists, this would be especially valuable to those who are not yet committed to a particular account of CAS. It would allow such philosophers a chance to judge the success or failure of distinct compatibilist accounts in relation to their own intuitive reactions to cases—for instance, one might think that a hard-line was taken when a soft-line was needed, or vice versa. I think it would be reasonable for one to be suspicious of compatibilists (and their respective views) who too frequently endorse a hard-line reply—often judging severely manipulated agents to be morally responsible—even when many other compatibilists are able to provide a soft-line reply that avoids such unsavory commitments. (Thus, while Haji is right to conclude that the 4-CA does not speak directly against the hard-liner such as Frankfurt, this reveals that a hard-line reply to the 4-CA could speak *indirectly* against the plausibility of that account; Haji, 24.)

[36] It leaves open the possibility that particularly clever manipulation arguments will provide a compatibilist with insight that will lead to the discovery of a new condition that results in a more adequate account of CAS. (One might think that McKenna's strategy would tend to cut off such a possibility prematurely.)

[37] It leaves the possibility of defeat wide open. (Haji might emphasize that his strategy leaves this possibility open, while McKenna's strategy sometimes appears to be recommending that compatibilists who have exhausted all other alternatives should try to achieve a stalemate by adopting a dogmatic, circling-of-the-wagons reply.)

One Last Note

[38] As a side note, it seems that Haji slightly misrepresents the content of hard incompatibilism (Haji, 3). While Pereboom rejects the truth of agent-causal libertarian theories for different reasons than event-causal compatibilist and libertarian theories (he thinks the former are contingently false while the latter theories are metaphysically impossible), I believe that it is his commitment to the falsity of *all* of these views at the actual world that makes him a *hard incompatibilist*.

[39] It is an unfortunate artifact of popular lingo that a libertarian view is committed to the truth of the libertarian account in the actual world, so that the theory is technically false if it is false in the actual world, while a compatibilist theory is technically false only if it is metaphysically impossible for a determined agent to be free and responsible. With this in mind, perhaps we should recognize different 'degrees of hardness' among hard incompatibilists: the hardest would be those who believe that free and responsible agency is metaphysically impossible, while the less hard incompatibilists would include all those who, like Pereboom, think that free and responsible agents are metaphysically possible, but that none happen to exist in the actual world.