Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin¹

MICHAEL MCKENNA

University of Arizona

In the pages of this journal, Derk Pereboom and I have debated the soundness of his Four-Case Manipulation Argument for incompatibilism (Pereboom, 2008; McKenna, 2008). In this paper, I develop further my hard-line compatibilist reply to him. My reply takes a hard-line insofar as it involves resisting the attractive intuitive judgment that a radically manipulated agent is not free and morally responsible.

1. Pereboom’s Four-Case Manipulation Argument

Pereboom’s argument is an instance of a form that builds upon a judgment about an example in which an agent is manipulated. The featured agent is manipulated into satisfying and acting from a Compatibilist-friendly Agential Structure (CAS), which provides conditions a compatibilist would take to be

¹ I presented slightly different drafts of this paper to the philosophy departments at University of Tennessee, California State University Los Angeles, Virginia Commonwealth University, to the Freedom Center and Philosophy Department at the University of Arizona, and to the 2012 Bowling Green State University Workshop in Applied Ethics and Policy on the topic of manipulation. I am grateful for the helpful comments offered by many in attendance at each session. Especially noteworthy were the remarks offered by E.J. Coffman, John Hardwig, David Palmer, David Reidy, and Clerk Shaw at the Tennessee session; Mark Balaguer, Talia Bettcher, Richard Dean, and David Pitt at the Los Angeles session; Gene Mills and Tony Ellis at the Virginia session; Mark Budolfson, Juan Comesana, Stew Cohen, Rachana Kamtekar, Shaun Nichols, David Owens, David Schmidt, and Hannah Tierney at the Arizona session; and Marcia Baron, Eric Cave, Christian Coons, Moti Gorin, Todd Long, David Shoemaker, Kevin Vallier, and Michael Weber at the BGSU session. John Martin Fischer, Todd Long, Alfred Mele, Carolina Sartorio, Patrick Todd, Brandon Warmke, and Jonathan Weinberg have offered excellent written comments, for which I am thankful. I have also profited from conversations with Terry Horgan and Jenann Ismael. I am especially indebted to Derk Pereboom for many rewarding and enjoyable conversations about the disagreement between us.
minimally sufficient for acting freely. Such an example used in the context of an argument for incompatibilism is meant to elicit the intuition that, due to the pervasiveness of the manipulation, the agent does not act freely and is not morally responsible for what she does. It is then claimed that any agent’s coming to be in the same psychic state through a deterministic process is no different in any relevant respect from the pertinent manner of manipulation. Hence, it is concluded that CAS is inadequate.

Pereboom attacks a version of CAS that is a conjunction of four influential contemporary compatibilist proposals: Ayer’s, Frankfurt’s, Fischer and Ravizza’s, and Wallace’s. For ease of discussion, let us refer to this conjunctive account as ‘CAS\*’. Given CAS*, the manipulated agent featured in Pereboom’s argument is not constrained by any irresistible desires and does not act out of character (Ayer, 1954); he acts upon first-order desires with which he identifies at a high-order (Frankfurt, 1971); he is responsive to reasons in a way that displays a stable, sane pattern (Fischer and Ravizza, 1998); and when he acts, he has the general capacity to regulate his behavior in light of moral reasons (Wallace, 1994). Although he does not say so explicitly, Pereboom clearly intends to target more than extant compatibilist proposals; all compatibilist comers are open to the same treatment. Whatever further compatibilist proposals one might come up with could just be conjoined to CAS*.

Pereboom does not work from just one manipulation case. Rather, he employs a series of them, three, in such a way that the first in the series involves a kind of manipulation that is very different from the kind of casual process that would be expected to unfold at a deterministic world of the sort that ours might (could?) be. He then marches through two more cases so that each case comes closer and closer to a case in which an agent acts as she does in a world at which determinism is true. In doing so, he employs a generalization strategy of treating like cases alike, and in transition between the cases, argues that there is no relevant difference between the modes of manipulation, so that if we are to treat the first case as one in which the manipulated agent is not free and responsible, that will carry over to the next in the series, and so forth, with the fourth and last merely being a case in which determinism is true.

Here is a truncated summary of each case:

Case 1: Professor Plum was created by a team of neuroscientists, who can manipulate him directly through radio-like technology, but he is as much like an ordinary human being as is possible, given his history. The scientists “locally” manipulate him to undertake a process of reasoning, directly producing his every state moment by moment, which leads to the killing of Ms. White for egoistic reasons. (Pereboom, 2001: 112–3)

The manipulation is such that Plum fully satisfies CAS* when he murders Ms. White. Of course, we are to have the intuition that, due to the extremity
of the manipulation, Plum does not act freely and is not morally responsible for killing Ms. White.

Case 2: Plum is like an ordinary human being, except he was created by a team of neuroscientists who, although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively egoistic, with the result that in the circumstances he is casually determined to undertake the process that results in his killing Ms. White. (Pereboom, 2001: 113–4)

Again, Plum satisfies CAS*. As for the time lag between the neuroscientists’ manipulations and Plum’s act, there is no relevant difference, Pereboom argues, between this case and Case 1 where the manipulation is direct and moment by moment. Treating like cases alike, we are to have the intuition that Plum does not act freely and is not morally responsible for killing Ms. White.

Case 3: Plum is an ordinary human being, except that he was determined by the rigorous training practices of his home and community so that he is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2). His training took place at too early an age for him to have had the ability to prevent or alter the practices that determined his character. In his current circumstances, Plum is thereby caused to undertake the… process… that results in his killing Ms. White. (Pereboom, 2001: 114)

Yet again, Plum satisfies CAS*. Case 3 is a common sort of case that we would find in the actual world. But the causal inputs, Pereboom reasons, are just less weird causes as in comparison with the sort highlighted in Case 2. Otherwise, there is no relevant difference. So, treating like cases alike, we are to have the intuition that Plum does not act freely and is not morally responsible for killing Ms. White.

Case 4: Physicalist determinism is true, and Plum is an ordinary human being, generated and raised under normal circumstances, who is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1-3). Plum’s killing of Ms. White comes about as a result of his undertaking the… [relevant] process. ([my braces], 2001: 115)

Once again, Plum satisfies CAS*. And since Case 4 differs from Case 3 merely by virtue of the fact the deterministic causes apply universally rather than locally as regards the details of Plum’s upbringing, there again seems to be no relevant difference between the cases. So, Pereboom argues, treating like cases alike, we should conclude that Plum in Case 4 does not act freely and is not morally responsible for killing Ms. White. Ergo, CAS* is refuted.
According to Pereboom, by way of his manipulation argument and his march from Cases 1 through 4, he is able to exploit Spinoza’s strategy of making salient the hidden causes of human actions so as to show how, once they are revealed, the illusion of control dissipates:

Men think themselves free, because they are conscious of their volitions and appetite, and do not think, even in their dreams, of the causes by which they are disposed to wanting and willing, because they are ignorant of [those causes]. (Spinoza, 1677/1985, v. 1, 440, as quoted in Pereboom, 2008: 161)

Seeing that being causally determined (Case 4) really is just like, in all relevant respects, being manipulated moment by moment by a team of neuroscientists (Case 1) helps crystallize the incompatibilist intuition that determinism destroys free and responsible agency.

2. A Hard-line Reply

Several critics have opted for a soft-line reply to Pereboom (e.g., Demetriou, 2010; Fischer, 2004; Haji, 1998, 2009; and Mele, 2005). A soft-line reply seeks some relevant difference between an agent, such as Plum in Case 1 or Case 2, manipulated into acting from a certain state, and an agent acting from a deterministic history (like Plum in Case 4). This allows the compatibilist to grant to Pereboom that, indeed, Plum in Case 1 or 2 is not free and responsible. Instead, the compatibilist can argue, there is some further feature of CAS that is lacking in the case of manipulation that is not lacking with relevant agents acting at a determined world. The trouble with this strategy, as I have argued elsewhere (2008), is that it is dialectically unstable. A compatibilist might very well be able to refute some specific formulation of Pereboom’s argument. But then it is open to Pereboom to revise his examples so that the feature of CAS alleged to be lacking in prior iterations of his argument is now present. This compatibilist short-game, it seems to me, eventually leads to a long-game victory for Pereboom—unless compatibilists can show that there is a principled reason why compatibilist sufficient conditions for freedom cannot be replicated by the design of a powerful being (a prospect I think it hopeless). Hence, I think it best for compatibilists to grant Pereboom that his cases satisfy what they require for CAS. And if his cases fall shy a bit, rather than call him out on this sort of technicality via a soft-line reply, they should just aid him by amending his case(s) so as to get right all that is required for CAS on their favored view. Having dispatched this preliminary work, they should then opt for a hard-line reply. This, at any rate, is how I have elected to take issue with Pereboom.

Turning to a hard-line reply, I assume that I must resist the contention that Plum in Case 1 does not act freely and is not morally responsible. To
achieve this result, in an earlier piece (2008), I first focused attention on agential and moral properties of Plum as displayed in Case 4. In this case, Plum is just as much like an ordinary moral agent as any one of us. His phenomenology is just as sophisticated. He is capable of feeling incredible remorse, or instead ambivalence about killing Ms. White, or alternatively, delicious pride. Furthermore, we can suppose that Plum had a rich history of moral development just like any psychologically healthy person who emerges from childhood into adulthood. And we can also allow that Plum is an agent who lives up to the kind of emotional complexity of the sort Strawsonians highlight. He is capable of resentment and gratitude, moral indignation and approbation, guilt and pride, and he is richly sensitive to the significance of these morally reactive emotions in his relations with others. He is, in this way, a full-fledged member of the moral community, one whose standing enables the kinds of interpersonal relations out of which adult moral life arises. The point of attending to these details is to help elicit the intuitions that are friendly to compatibilists—to bring forth the sense that a determined agent of the sort Plum is in Case 4 is a richly complex agent just like any person we might come across.

Second, I sought to clarify the initial attitude both compatibilists and incompatibilists are entitled to adopt toward Plum in Case 4. Naturally, the compatibilist is not entitled to insist that it be granted that Plum in Case 4 does act freely and is morally responsible. That is precisely what is in dispute in considering the manipulation argument currently on the table. Nor, for this reason, should the compatibilist expect that the intuitions that she seeks to elicit in Case 4—by focusing on salient agential and moral properties—speak commandingly in favor of compatibilism. All the compatibilist can expect is that an open-minded inquirer, one who is undecided and so agnostic by the prospect that determinism rules out freedom and responsibility, could be moved to see the plausibility of compatibilism. But equally to the point, incompatibilists cannot assume at the outset that Plum in Case 4 is not free and is not morally responsible, nor ought they to expect that the initial intuition about Plum in Case 4 is that, clearly, he does not act freely and is not morally responsible. Again, that is what is in dispute, and assuming this at the outset would be patently question-begging. Thus, the stance towards Case 4 that an incompatibilist such as Pereboom is dialectically obligated to take in executing his argument is the same one that the compatibilist must take. Both must allow that in Case 4, it is an open question whether Plum acts freely and is morally responsible for killing Ms. White.

Third, armed with the two preceding considerations, I have attempted to turn Pereboom’s strategy against him. Pereboom begins with his Case 1 and, seeking to elicit an incompatibilist intuition in response to the disturbing manipulation featured in it, marches through to his Case 4, thereby
drawing an incompatibilist conclusion. I grant that Pereboom does gain some intuitive advantage in favor of incompatibilism by calling attention in Spinozistic fashion to hidden causes revealed in Case 1 and transferring through to Case 4. That, I acknowledge, has to be weighed on the scales when evaluating the overall strength of his argument. Nevertheless, I propose that the compatibilist move in the opposite direction, and that this too must be weighed on the scales.

Begin with Pereboom’s Case 4. By calling attention to the salient agential and moral properties of Plum in Case 4, I have sought to elicit the intuition amongst an audience of undecided, open-minded inquirers that it is not clearly the case that determinism rules out free will and moral responsibility. It is, at least, a plausible contention. Since, by Pereboom’s own lights, Case 3 differs in no relevant respect from Case 4, and since we are also instructed to treat like cases alike, the compatibilist should argue that we should treat Case 3 just as we treat Case 4. Likewise we should treat Case 2 no differently than we treat Case 3, just as Pereboom contends, and also, finally, Case 1 no differently than we treat Case 2. Hence, we get the result that we should treat Case 1 no differently than we treat Case 4.

To clarify: how, exactly, should we treat Case 1, according to this hard-line compatibilist strategy? We should transfer the initial undecided open-minded attitude of agnosticism that perhaps Plum in Case 4 does act freely and is morally responsible. That is, given this strategy, and the initial dialectical attitude the compatibilist is entitled to in Case 4, the compatibilist is in no position to conclude that in Case 1 Plum does act freely and is morally responsible. She’s only entitled to conclude that it is not clear that Plum does not act freely and is not morally responsible. And she is entitled to conclude this much by way of carrying over the intuitive weight regarding Case 4 that she seeks initially to elicit from an undecided audience by inviting them to focus carefully on the rich agential and moral properties that Plum in Case 4 has—all of which, it can be granted, carry over just as much to Plum in Cases 3, 2, and 1.2

Of course, it will not go unnoticed that my proposed compatibilist hard-line response does not require taking the bold position of maintaining that in Case 1 Plum does act freely and is morally responsible. Some will likely regard this as a copout. I do not think it is. The dialectical burden of the compatibilist who adopts the hard-line in response to Pereboom is simply to discredit the first premise in this manipulation argument. This is achieved merely by casting sufficient doubt on the truth of the premise. It’s the

---

2 To heighten the effect of my strategy, I added two cases to Pereboom’s four cases so as to make the transition from a mere determined agent to Plum in Pereboom’s Case 1 even smoother. I’ll set that detail aside here, since nothing in my current response to Pereboom will turn on it.
incompatibilist who has offered the argument, and it’s he who is claiming that intuitions about a case like Case 1 will speak commandingly in favor of the proposition that the featured agent does not act freely and is not morally responsible. It’s not the compatibilists’ burden in this context to convince fair and open-minded inquirers that clearly an agent manipulated the way Plum is in Case 1 does act freely and is morally responsible. This crowd—the audience of fair and open-minded inquirers—was not even expected to have that intuition about Case 4 antecedent to reflecting on the more outrageous cases like Cases 1 and 2.

To help cement this hard-line compatibilist response to Pereboom, I took one further step to lessen the jarring counterintuitive suggestion that, perhaps, an agent like Plum in Case 1 does act freely and is morally responsible. The central idea is that there are actual cases which are in many respects just like the imaginary outlandish manipulation cases featured in an argument like Pereboom’s, and these cases are not that uncommon. Here, our intuitions do not seem to favor an incompatibilist diagnosis. As Nomy Arpaly has thoughtfully pointed out when reflecting on the nature of different manipulation cases, there are all sorts of real life cases where persons undergo spectacular changes as a result of things like religious conversions, the birth of a child, one’s hormones drying up, and so on (Arpaly, 2006: 112–3). Indeed, there’s no end to the number of naturally occurring real-life cases of “manipulation” in which massive and unexpected alterations disrupt people’s lives in ways that dramatically reconfigure their psychic constitutions: people suffer traumatic accidents; have a loved one die in their arms unexpectedly; are crushed during their youths by the weight of violent parents. Yet, if these changes leave the adult person otherwise, sane, rational and stable, in ways that allow satisfaction of pertinent compatibilist conditions on free action, very few are inclined to think that such “manipulation” cases undermine a person’s free agency and her responsibility. They are just regarded as the (often) tragic contingencies out of which these people manage to fashion their moral personalities and thereby come to be the kinds of agents they are.

3. Pereboom’s Resistance to the Hard-line Reply

To resist my hard-line reply, Pereboom challenges my contention that an initial agnosticism towards Case 4 should transfer through to Case 1. Rather, he argues, if we get clear on the dialectically appropriate state that should be expected of those whose intuitions ought to be informative, we shall see that cases like his Case 1 and Case 2 should still have the effect of recommending an incompatibilist diagnosis as regards Case 4.

One point of agreement between Pereboom and me has to do with those whose intuitions are not the target of a manipulation argument
There are some who, upon hearing about the puzzle regarding freedom and determinism, are, so to speak, natural incompatibilists; their untutored intuitions are such that they need not be prompted by reflection upon manipulation cases and a corresponding argument, or for that matter any other sort of incompatibilist argument, to adopt an incompatibilist posture. There are others who, upon learning of the puzzle about free will, are instead natural compatibilists; their untutored reactions are to think that determinism does not pose much of a threat to freedom and responsibility at all, and they do not need to be prompted by further arguments on behalf of compatibilism. In either of these groups—those of the natural incompatibilists as well as the natural compatibilists—intuitions about responses to manipulation cases will not prove fruitful for settling the dispute between compatibilists and incompatibilists. Natural incompatibilists can be expected to respond to pertinent manipulation cases in ways that predictably confirm their predisposition. And natural compatibilists can be expected to do likewise. (Or, instead, if natural compatibilists do respond differently to learning of relevant manipulation cases, it is likely that they will be inclined to retain the conviction that there must be some relevant difference between determinism and such manipulation, even if there is difficulty stating what that comes to.) So the turf over which incompatibilists and compatibilists are doing battle with respect to manipulation arguments like Pereboom’s four-case argument is the turf of the undecided. But how are we to understand the appropriate initial epistemic stance of this undecided group in relation to the stance they ought to adopt after reflecting upon cases like Pereboom’s Case 1 and Case 2?

According to Pereboom, my implicit answer to this question shows where I have gone wrong, and why it is that he is entitled to claim that cases like his Case 1 and 2 can be counted on to generate in the appropriate audience an incompatibilist conclusion. Here is how Pereboom formulates my position:

[McKenna’s] central idea is that whatever attitude it is rational initially to have about Plum’s responsibility in the ordinary deterministic examples (my [Pereboom’s] case 4…) transfers to the manipulation cases, since, after all, the incompatibilist has set up the cases so that there are no responsibility-relevant differences among them. Because it is at the outset rational for us to have an agnostic attitude about the claim that Plum is morally responsible in the ordinary deterministic examples, the absence of relevant differences allows the rational agnosticism to transfer unimpeded to the manipulation cases, thereby depriving them of counting in favor of incompatibilism. (2008: [my braces] 160)

Pereboom then attributes to me the following proposition as it applies to the initial attitude one ought to adopt towards his Case 4:
It is this attitude, (M), that I contend is a rational one to adopt as transferring to cases 3 through 1. My mistake, according to Pereboom, is in failing to appreciate how it is that cases like his Case 1 and Case 2 could count as clarifying considerations that ought to lead the targeted audience to reconsider Case 4 and reject (M) in favor of an incompatibilist diagnosis.

To make his point perspicuous, Pereboom distinguishes between different “undecided” stances. One stance is the stance of the confirmed agnostic, who, upon initial reflection as regards the puzzle about freedom and determinism, is resolute in her agnosticism and believes that no further enquiry could settle the matter (162). As Pereboom points out, this audience could be expected to transfer (M) from a case like Case 4 through to Case 1; this is what they rationally ought to do given their initial attitude. If this were the dialectically appropriate audience to target, Pereboom, notes, then my reply to him would be warranted (163). But Pereboom rightly rejects this as the appropriate initial attitude. According to him, the appropriate approach:

…affirms that determinism provides a reason for giving up the responsibility assumption, but claims that so far the issue has not been settled. Its advocate would say about an ordinary case of an immoral action, in which it is specified that the action results from a causally deterministic process that traces back beyond the agent’s control, that it is now in question whether the agent is morally responsible. Call this the neutral inquiring response. By this response it is initially epistemically rational not to believe that that the agent in an ordinary deterministic example is morally responsible, and not to believe that he is not morally responsible, but to be open to clarifying considerations that would make one or the other of these beliefs rational. (162)

According to Pereboom, the appropriate initial response towards his Case 4 by members of the pertinent target audience is the neutral inquiring response (NIR). But what kind of “clarifying considerations” would be of use for such an audience? Pereboom writes:

…adducing an analogy in which one’s intuitions are clearer might itself count as the relevant sort of clarifying consideration. In a situation in which the neutral inquiring response to an ordinary determined agent is at first epistemically rational, it may be that an analogous manipulation case functions as a clarifying consideration that makes rational the belief that the ordinary determined agent is not morally responsible. (162)

Of course, as Pereboom sees it, intuitions about a case like Case 1 or Case 2 are clearly friendly to incompatibilism, and these cases provide the right sort of analogy to Case 4 to help clarify the rationality of one’s beliefs.
So, in short, according to Pereboom, my contention that (M) transfers through from Cases 4 through 1 rests on the fallacious assumption that the proper initial attitude towards Case 4 is that of the confirmed agnostic, not that of the open-minded and undecided neutral inquirer. Once the appropriate initial attitude is clarified, it can be seen that (M) will not—and, more importantly, rationally ought not—comfortably transfer through from Case 4 to Case 1. It ought rather to collapse under the weight of the intuitive force of cases like Case 1 and Case 2 as these will clarify what really is going on, even in an ordinary case like Case 4.

Building upon the preceding indictment, Pereboom entertains the possibility that a compatibilist might wish to contend that the neutral inquiring response initially adopted toward Case 4 ought to remain the appropriate attitude to adopt even upon confronting cases like his Case 1 and 2. If so, these cases would not after all count as clarifying considerations on behalf of incompatibilism (163–4). This could then be used by the compatibilist to place in check Pereboom’s reply to me. But here Pereboom correctly points out that this manner of resisting him would commit a compatibilist to the implausible claim that even initially upon facing a case like Case 1 or Case 2, there is no preliminary intuition of non-freedom and non-responsibility that such cases can be expected to elicit (164). This, he points out, would amount to a departure from my own position. Indeed, it would. How so?
The strategy I adopted of transferring one’s response toward Case 4 through to Case 1 was meant to function as a tonic to the expected initial reaction to Case 1 that, at least provisionally, an agent so manipulated is, intuitively, thought to be such that she does not act freely and is not morally responsible.

4. A Hard-liner Takes It on the Chin
Pereboom’s fair-minded response offers a constructive and insightful way to move the debate further along. But I do not think it is adequate to block my hard-line reply to his argument. I am in complete agreement with Pereboom as regards the initial dialectical stance that both compatibilists and incompatibilists should adopt when approaching his manipulation argument. Both compatibilists and incompatibilists need to suppose that the neutral inquiring response toward ordinary determined agents (of the sort featured in Pereboom’s Case 4) is initially epistemically rational. However, from this supposition, it would be wrong to infer that, short of admitting defeat, it is not open to a hard-liner to grant that manipulation cases like Pereboom’s Case 1 and Case 2 can count as clarifying considerations speaking on behalf of incompatibilism. Thinking this concession off limits, one might then assume that a hard-liner is fallaciously driven to adopt the stance of the confirmed agnostic. Otherwise, it might be thought, how can the
hard-liner argue that the initial attitude (M) does after all transfer from Case 4 to Case 1? Won’t the effects of Case 1 and Case 2 undermine (M)?

Here is the key inference that I wish to resist: If manipulation cases like Pereboom’s Case 1 and Case 2 provide intuitive weight in favor of a judgment of non-freedom and non-responsibility, then they count as clarifying considerations that are sufficiently strong that they ought rationally to move one off of a neutral inquiring stance as regards a case like his Case 4. It is this assumption that I wish to challenge. I grant that cases like Pereboom’s Case 1 and Case 2 do initially and presumptively weigh in the direction of judging that the Plums in each of them do not act freely and are not morally responsible. And I grant that, to this extent, they count as incompatibilist-friendly clarifying considerations as regards the initial neutral inquiring response toward Case 4. In making these concessions, I am, so to speak, taking it on the chin. Pereboom’s argument is a powerful one, and it would be nothing shy of intellectually dishonest not to recognize it as such. But what I wish to challenge is that his Cases 1 and 2 are sufficiently compelling to move one fully off the neutral inquiring response. Speaking in terms of degrees of credence, cases like Pereboom’s Case 1 and 2 might help move the neutral inquirer to raise her credence somewhat in favor of a judgment of incompatibilism as regards Case 4, and to that extent be clarifying, but, I shall now argue, there is reason to believe that such manipulation cases ought not be regarded as decisive. To do so, I offer the following three points.

First, Pereboom wishes to make clear that cases like his Case 1 and Case 2 count as clarifying considerations that ought to move relevant parties in the direction of an incompatibilist diagnosis. But equally to the point, consider those whose initial attitude toward a case like Case 1 or Case 2 is “Clearly, certainly, obviously, such an agent is not free and is not morally responsible!” Calling attention to the way in which these outrageously manipulated agents could be just like a normal person any one of us might meet (a Case 4), but with very strange or unexpected causal strings, also gets to count as a clarifying consideration. And this should provide some reason for those who would adopt a neutral inquiring response toward a case like Case 4 to at least consider looking askance at their initial intuitive reaction to such cases as Case 1 and Case 2. It is for this reason that, in advancing a hard-line reply, I have argued that the compatibilist needs to fill out the manipulation case as clearly as possible, so as to make clear that the candidate for manipulation is not made out to appear no more than, say, a puppet or a zombie—a mere comic book sketch of a full-blooded person.

---

3 To avoid any misimpressions, I am not attributing this inference to Pereboom. It’s just that, in making clear how I distance myself from it, I can bring into relief how it is that a hard-liner like me should resist Pereboom.
The modes of manipulation have to allow for an agent whose complexity and psychic depth are no different from any one of us. And, I contend, in advancing a hard-line reply, the compatibilist needs to make this vividly clear to her target audience of neutral and open-minded inquirers.

Second, while cases like Pereboom’s Case 1 and Case 2 are to be placed on the scales in weighing the intuitive force of incompatibilism, so too should the kinds of cases to which Arpaly has called attention. These cases, I have noted, are unfriendly to an incompatibilist diagnosis. And they too get to count as clarifying considerations—considerations that, I say, counterbalance at least to some degree the force of cases like Case 1 and Case 2. What’s especially noteworthy about these kinds of cases is that they are all akin to Pereboom’s Case 3, wherein Plum is simply raised by parents in an environment that has powerful influence on Plum’s values as he emerges into adulthood. In my earlier piece, to bring a case like this to life, I drew upon the example of Ann, a woman whose early childhood was shaped by the protracted death of a parent suffering through chemotherapy when that form of treatment was still in its infancy, when it was barbaric (2008: 256–7). Ann, I explained, came to value life in a certain way from those early experiences—as something that should not be squandered, that should be lived to its fullest, with no promise of a long future or a lovely afterlife. And so her life was shaped. Ann, I noted, did not regard these facts about her history as impediments to her freedom, but instead as sources of it, as conditions of her agency and sources of her autonomy. The case of Ann, I proposed, just is a real life “manipulation” case, but for the fact that it was not intentionally brought about by another agent for the purpose of so shaping Ann. For the most part, the natural, intuitive response to these kinds of cases is to persist in regarding such agents as free and responsible.

In his reply, Pereboom thoughtfully countenances my case of Ann (Pereboom, 2008: 167). He notes that he has available the conceptual resources to acknowledge the value of Ann’s life and its direction despite his disinclination to think of her as rendered free and responsible by her history. Her virtuous achievement, Pereboom remarks, can be enjoyed and celebrated (167). Furthermore, Pereboom is careful to acknowledge that cases like the Ann case do count in favor of a compatibilist diagnosis and against his own position. Indeed, he regards it as a cost of his position that it is not in keeping with the likely widespread response that a person like Ann is free and responsible with respect to her actions issuing from the relevant values. Pereboom’s remarks here strike me as especially charitable and fair-minded. But in reply, I would note that it is illuminating that he seeks alternative means to account for cases like the Ann case—a case very much in the family of cases like his Case 3. Why oughtn’t the case of Ann just be treated as a further clarifying consideration as regards the neutral inquiring response, a clarifying consideration that, in assessing (M), competes for credence with responses to cases like his Case 1 and Case 2? In fact, one
might wonder whether cases like the case of Ann should be afforded a greater degree of significance in this dispute between compatibilists and incompatibilists. It is to this point that I shall now turn.

Third, it is worth examining the relative status of intuitions about cases like Pereboom’s Case 1 and Case 2, as in comparison with intuitions about cases like the Arpaly cases and my case of Ann. It is one thing to reflect upon how those untutored subjects who are open-mined, neutral inquirers both would and ought rationally to respond to all of these cases: What grounds do these cases offer as a basis for moving this audience either off of (M) and in the direction of incompatibilism, or instead more confidently toward (M) and thus away from incompatibilism? But it is another for philosophers engaged in the free will dispute to assign these different responses from the pertinent audience their relative overall significance in assessing an argument like Pereboom’s four-case argument: How ought we as philosophers assess what the crowd of neutral inquirers both would and ought rationally to respond?

I propose that philosophers assessing manipulation arguments should assign responses from the neutral inquiring audience to cases like my case of Ann and the Arpaly cases a greater degree of clout as in comparison with cases like Pereboom’s Case 1 and Case 2. Grant that the audience of neutral inquirers both would respond and rationally ought to respond to Case 1 and Case 2 by ratcheting down their credence toward (M) in favor of an incompatibilist conclusion. And grant that this same audience both would respond and rationally ought to respond to cases like my case of Ann and the Arpaly cases by ratcheting up their credence toward (M) and thereby move way from an incompatibilist conclusion. Suppose, furthermore, just for argument’s sake, that the net overall effect for this audience both would be and rationally ought to be a lowering of credence toward (M) in favor of incompatibilism. On its face, this would seem to favor Pereboom’s conclusion. But should we as philosophers just read off directly from these results our philosophical conclusions? I don’t think so. I think there are good theoretical reasons to favor at least somewhat responses from neutral inquirers to cases like my case of Ann and the Arpaly cases as in contrast with cases like Pereboom’s Case 1 and Case 2—and these theoretical reasons are not reasons to which we should expected an untutored audience to be especially sensitive.

This is a point I mentioned for a slightly different purpose in my original reply to Pereboom (McKenna, 2008: 157).4 It is reasonable to suppose that

---

4 To clarify, originally (in my 2008) I was not interested, as I am here, in how the effects of these different cases on the relevant target audience ought to bear on philosophers’ assessments of a proposition (M). Presently, I am interested in distinguishing how we philosophers ought to assess the force of these different cases on (M) as in comparison with—and in light of—how members of an untutored audience would and ought to do so.
our intuitions have evolved along with our ordinary practices. Here, what I mean by intuitions are judgments in response to concrete cases expressing how pertinent concepts apply to them. These judgments, as Pereboom and I are using them, are meant to help settle the dispute between us. It is my contention that the further away from ordinary contexts the application of these concepts are, the less reliable we should take them to be. There’s a natural explanation for this. Our training for these concepts involved applications to contexts structured by our natural surroundings and our entire form of life. The more we move away from these, the less mooring we theorists have to feel confident that ordinary users are indeed applying our concepts properly. This is not to say that the application of our concepts in very outlandish contexts should be assigned no significance. It’s just to say that we as philosophers have good reason, as a general policy, to assign a higher degree of credence to those untutored judgments closer to ordinary contexts—regardless of the degree of confidence untutored subjects might assign to their judgments about outlandish cases when prompted.

The preceding point needs to be dispatched with caution. After all, as Herman Cappelen has pointed out, mere outlandishness of a case is consistent with applying a concept to it with a high degree of reliability. Cappelen, for instance, gives the following example (2012: 226):

There are two pink elephants in his office and another enters. How many pink elephants are in his office?

Here, our application of the concept of three is not any less reliable simply because the case is outlandish. Of course, Cappelen is correct about this. But I would offer three qualifications to the point I mean to press: First, note that in Cappelen’s example the role played by the pink elephants can be factored out of the thought experiment. It’s about there being two things of a certain kind in a room and what judgment we should draw if we add one more. While the metric of proximity to the actual world suggest that the case is far away, the metric at issue here is with conditions of proximity for familiar application of the concept. We are very comfortable, and we operate in nearby psychological space, when we add one thing, whatever kind of thing it is, to two things. This is so even if it is a relatively distant world we have to travel to for there to exist such things. But when we attempt to apply a concept like moral responsibility to beings whose origins and causal strings are quite divergent, we are not in nearby psychological space. Our training, so to speak, for acquiring and competently applying the concept in these ways did not encompass such contexts. Second, my contention regarding the

---

5 I would like to thank an anonymous referee for drawing my attention to this challenge and to Cappelen’s useful example.
status of intuitions about these sorts of bizarre cases is not a strongly skeptical thesis. I am not arguing, as others apparently are, that our judgments about these sorts of bizarre cases are corrupted and so should be assigned no credence—should not be regarded as evidence at all (e.g., Machery, 2011: 202–3). I am only essaying for the relatively modest and philosophically conservative thesis that their force should not be taken to trump our intuitive judgments about closer-to-life cases unless we have very good reason to think that either the closer-to-life cases really are corrupted for some reason, or instead the bizarre cases acquire a special status that warrants us in thinking they are more useful or provide greater clarity than our judgments about more mundane cases. Third, and finally, note that my claim here is domain-specific at least in this respect: if we had some special reason to think our reaction to cases like the Arpaly cases were benighted or that our reactions to the bizarre cases like Pereboom’s Case 1 offered extra illumination that the more mundane cases lacked, then we would have a defeater to my modest claim. But we have none of that—or so I say. All I am claiming is that other-things-being-equal, intuitive reactions to closer-to-home cases offer a higher degree of reliability given that our conceptual training and modes of performance in these contexts is cultivated and honed in ways that they are not when we apply them in the psychologically bizarre cases.

Cases like the case of Plum 1, for instance, seem so alien, and it can seem very easy to apply very confidently the concept of moral responsibility to the case. But as we have not evolved in a context in which it might be expected that some or all of us are caused to act as we do by teams of neuroscientists, rather than, say, our own DNA and our local environments, it is understandable that such contexts would generate a strong negative reaction. However, it is just not clear what our concept of responsibility would look like if instead it evolved in the context of that very different sort of life. Hence, as I remarked earlier (2008: 157), if incompatibilists such as Pereboom wish to barter dialectically in the currency of intuitions that are lifted out of the contexts that are their natural homes, then when they generate intuitions from wildly deviant cases, it is incumbent upon them to hold themselves to very high standards indeed.

In my opinion, once we secure these standards, compatibilists ought to concede to Pereboom that, for a neutral inquiring audience, case like his Case 1 and Case 2 place some degree-of-credence-lowering pressure on (M). Furthermore, it can even be granted that this credence-lowering pressure has the net effect of moving neutral inquirers to some degree in the direction of incompatibilism when counterbalanced with the credence-raising pressure of cases like my case of Ann or the Arpaly cases. But in light of the theoretical worries raised here, this should not be regarded as decisive. Even if, amongst neutral inquirers, intuitions about cases like Case 1 and Case 2 weigh more heavily than intuitions about my case of Ann and
the Arpaly cases, they do not weigh heavily enough for the incompatibilists to claim victory. Maybe (M) is weakened, but it is not refuted.

Given the three preceding points, I conclude that Pereboom has not offered sufficient reason to believe that amongst an audience of neutral inquirers, his clarifying cases Case 1 and Case 2 give this audience sufficient reason to contend that (M) is false. And to the extent that his clarifying considerations give this audience good reason to lower somewhat their credence toward (M), there are theoretical reasons not to regard this as decisive in settling the dispute between compatibilists and incompatibilists. Admittedly, given my strategy of transferring the initially rational attitude it is appropriate to have toward Plum in Case 4, and given the concessions I have made here, (M) does not arrive at Case 1 untaurished. It’s a bumpy ride, no doubt. But as a compatibilist pressing a hard-line reply, I claim that (M) does after all arrive at Case 1, even if it arrives weaker than it was when it first departed.6

5. Conclusion

With all that has been said, doesn’t it count against compatibilism that an attitude such as (M) is weakened by manipulation cases like Pereboom’s? If such cases ought to move fair and open-minded neutral inquirers even somewhat in the direction of incompatibilism, doesn’t this amount to a reason speaking in favor of incompatibilism? It seems to me that the honest answer is ‘yes’; the fact that a case like Pereboom’s Case 1 ought to move neutral inquirers even somewhat in the direction of incompatibilism is a reason in favor of incompatibilism. It’s just that it’s not a decisive reason. Perhaps this will be regarded by some compatibilists as granting far too much. Some compatibilists (e.g., Dennett, 1984) tend to write as if there is not anything in the least counter-intuitive about compatibilism—that intuition unqualifiedly favors the compatibilist position. I disagree. Compatibilism is not an intuitive thesis through and through. It is an odd view that does after all bump up against some of our intuitions. Anyone’s experience teaching the free will problem in an introductory philosophy class confirms the point. It is only to be expected that at points compatibilism winds up looking a bit embarrassed when exposed in nothing but its underwear. Cases like Case 1, I say, show compatibilism for what it is quite vividly—just as, Pereboom contends, Spinoza would have it. But it is a mistake to draw a conclusion against compatibilism for this sort of reason. Incompatib-

6 This shows how I am inclined to reply to an especially impressive incompatibilist move recently made by Patrick Todd (2011). Todd regards it as conclusive in favor of incompatibilism that intuitive responses to manipulation cases mitigate blame, and by extension, assumptions about blameworthiness. Here, I am willing to make the concession regarding mitigation but resist the claim that it is decisive. (Space does not permit a proper rely to Todd here, although I intend to develop one elsewhere.)
ilism exposed in nothing but its tidy whites is no prettier; it too leads to counter-intuitive results. So, in the end, we need to weigh costs.

It is a hard truth that, as Harry Frankfurt remarked (2002: 28), we are inevitably fashioned and sustained by conditions over which we have no control. Compatibilists see it as a truth we can live with while persisting in regarding ourselves and others as free and responsible—even in the basic desert-entailing sense. Incompatibilists like Pereboom do not. Manipulation cases like Pereboom’s Case 1 and Case 2 make it harder for the compatibilist to stomach that hard truth, but I have argued that they do not provide decisive grounds for giving up the compatibilist position altogether. They may very well lead us to pine for an existence in which we do fashion and sustain ourselves, at least to some extent, by conditions over which we do ultimately have control, so that there is no way that we could be like Plum in Case 1. But it is unclear whether or not this notion of being an ultimate creator and sustainer of oneself is even a coherent metaphysical possibility. Even if it is a metaphysical possibility, but one beyond the reach of limited creatures like ourselves, why should we conclude that, falling shy of it, nothing short of that is good enough?

References


