

Summer 8-11-2015

# Why Pereboom's Four-Case Manipulation Argument is Manipulative

Jay Spitzley

Follow this and additional works at: [https://scholarworks.gsu.edu/philosophy\\_theses](https://scholarworks.gsu.edu/philosophy_theses)

---

## Recommended Citation

Spitzley, Jay, "Why Pereboom's Four-Case Manipulation Argument is Manipulative." Thesis, Georgia State University, 2015.  
[https://scholarworks.gsu.edu/philosophy\\_theses/172](https://scholarworks.gsu.edu/philosophy_theses/172)

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

# WHY PEREBOOM'S FOUR-CASE MANIPULATION ARGUMENT IS MANIPULATIVE

By

JAY SPITZLEY

Under the Direction of Eddy Nahmias, PhD

## ABSTRACT

Research suggests that intuitions about thought experiments are vulnerable to a wide array of seemingly irrelevant factors. I argue that when arguments hinge on the use of intuitions about thought experiments, research on the subtle factors that affect intuitions must be taken seriously. To demonstrate how failing to consider such psychological influences can undermine an argument, I discuss Pereboom's four-case manipulation argument. I argue that by failing to consider the impact of subtle psychological influences such as order effects, Pereboom likely mis-identifies what really leads us to have the intuitions that we have about his cases, and this in turn undermines his argument for incompatibilism. Last, I consider objections and discuss how to empirically test my hypothesis.

INDEX WORDS: intuitions, moral responsibility, free will, manipulation, Derk Pereboom

WHY PEREBOOM'S FOUR-CASE MANIPULATION ARGUMENT IS MANIPULATIVE

By

JAY SPITZLEY

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Masters of Arts

In the College of Arts and Sciences

Georgia State University

2015

Copyright by  
Jay Spitzley  
2015

WHY PEREBOOM'S FOUR-CASE MANIPULATION ARGUMENT IS MANIPULATIVE

By

JAY SPITZLEY

Committee Chair: Eddy Nahmias

Committee: Nicole Vincent

Neil Van Leeuwen

Electronic Version Approved:

Office of Graduate Studies  
College of Arts and Sciences  
Georgia State University  
May 2015

**TABLE OF CONTENTS**

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>2</b>	<b>MANIPULATION CASES AND ARGUMENTS.....</b>	<b>3</b>
<b>3</b>	<b>PEREBOOM’S FOUR-CASE MANIPULATION ARGUMENT .....</b>	<b>12</b>
<b>4</b>	<b>ORDER EFFECTS AS AN ALTERNATIVE EXPLANATION .....</b>	<b>18</b>
<b>4.1</b>	<b><i>Agency-Detection Mechanism .....</i></b>	<b>20</b>
<b>4.2</b>	<b><i>Intent.....</i></b>	<b>22</b>
<b>4.3</b>	<b><i>Emotional Responses .....</i></b>	<b>24</b>
<b>5</b>	<b>OBJECTIONS AND EVIDENCE.....</b>	<b>28</b>
<b>5.1</b>	<b><i>Order Effects Are Intended.....</i></b>	<b>28</b>
<b>5.2</b>	<b><i>Explaining Intuitions is Unimportant .....</i></b>	<b>30</b>
<b>5.3</b>	<b><i>Providing Evidence for the Best Explanation .....</i></b>	<b>34</b>
<b>6</b>	<b>CONCLUSION .....</b>	<b>39</b>
	<b>REFERENCES.....</b>	<b>42</b>

## 1 INTRODUCTION

Philosophers commonly use thought experiments and hypothetical cases to bolster their arguments. While many believe that intuitions about these cases are reliable, recent empirical research shows that certain features of some hypothetical scenarios may lead our intuitions to be unreliable. For instance, philosophers who contribute to the “negative” program in experimental philosophy have demonstrated that intuitions vary according to ethnicity (Weinberg et al. 2001), gender (Buckwalter and Stich 2011), and linguistic background (Vaesen et al. 2013).<sup>1</sup> Further research shows that moral judgments are significantly influenced by trivial and irrelevant factors of hypothetical cases, such as the order in which information is presented (Weigmann et al. 2012; Schwitzgebel and Cushman 2012), the way in which information is worded (Petrinovich and O’Neill 1996), the emotional state of the reader (King and Hicks 2011; He et al. 2013; Guiseppe et al. 2012), and even the smell of Lysol (Tobia et al. 2013). Furthermore, there is also overwhelming evidence that humans are unaware that such factors influence their judgments (King and Hicks 2011; Mlodinow 2012; Li et al. 2008), and even that philosophers are susceptible to unconscious psychological influences (Schwitzgebel and Cushman 2012; Tobia 2013). Despite these complications, philosophers frequently assume intuitions regarding thought experiments are driven by reliable processes and relevant features of the thought experiments. Because intuitions seem to vary across demographics and seem to be significantly influenced by irrelevant features, some argue that intuitions about thought experiments should not count as evidence in support of philosophical views (Weinberg 2008; Sinnott-Armstrong 2008).

---

<sup>1</sup> Experimental philosophy’s “negative program,” generally seeks to challenge the usefulness of appealing to intuitions as a legitimate philosophical methodology for uncovering justified beliefs (Alexander, Mallon, and Weinberg 2014).

While I do not address the complications which arise from intuitions varying across demographics (in part, because some these complications may not be as concerning as they first appear (Adelberg, Thompson, and Nahmias forthcoming)), I *do* argue that there are good reasons to maintain that some intuitions that are relied upon for philosophical argumentation are significantly influenced by seemingly irrelevant factors. To be clear, I will not be arguing that it is never legitimate to rely upon intuitions (Weinberg 2008) or even that moral heuristics distort our intuitions to the point that we have good reason not to trust them in certain situations (Sinnott-Armstrong, Young, and Cushman 2009). Instead, I argue that we need to consider whether unconscious psychological influences that affect our intuitions and moral judgments may undermine the use of intuitions in some philosophical arguments. Given both the influence that these factors have on intuitions about thought experiments and how important it is that intuitions about thought experiments are tracking the right kinds of features, I argue there are some instances where intuitions can be better explained by these psychological features of which we are not aware than by the relevant features of thought experiments. Therefore, when presenting an argument that relies on an explanation for what features of a case motivate intuitions about that case, one must acknowledge these features as alternative explanations for intuitions. Failure to consider psychological influences, some of which may be entirely unconscious, as alternative explanations for what drives intuitions would be a methodological error and could undermine philosophical arguments.

To demonstrate how failing to take these psychological influences, such as order effects, into account when doing philosophy can undermine one's argument, I discuss Derk Pereboom's (2014, 2002) four-case manipulation argument for incompatibilism. I begin in section 2 by discussing manipulation arguments in general and I describe what conditions they must meet in



order to be successful. In section 3, I introduce Pereboom's four-case manipulation argument and draw attention to how important it is for Pereboom that he offers the best explanation for what generates the intuitions that we have about his four cases. In section 4, I introduce unconscious psychological factors that likely drive intuitions about Pereboom's manipulation cases. Because these factors significantly influence intuitions and moral judgments and yet Pereboom fails to address them in his explanation, Pereboom's manipulation argument fails to meet all of the conditions required in order for a manipulation argument to succeed as an argument for incompatibilism. Therefore, by neglecting to take into account order effects and unconscious influences that drive order effects, Pereboom's argument for incompatibilism is open to a serious objection and is likely undermined. In section 5, I consider objections to my argument and discuss how we could design empirical tests to determine whether my hypothesis is correct. While my argument currently has only indirect empirical support, by bringing attention to the wealth of empirical research that suggests there are other judgment-influencing factors than the ones Pereboom addresses, the burden of proof falls on Pereboom to demonstrate that his explanation for what drives intuitions in response to his thought experiments is indeed the best explanation.

## **2 MANIPULATION CASES AND ARGUMENTS**

Manipulation cases, generally, are introduced to serve as counterexamples to compatibilism, where compatibilism is the thesis that determinism does not necessarily rule out free will and moral responsibility. While definitions for terms like 'free will' and 'moral responsibility' are far from universally agreed upon, for the purpose of this thesis, I will adopt Derk Pereboom's (2014) definitions of these terms. Pereboom adopts Mele's (2006) notion of free will and takes it "to refer to the strongest sort of control in action required for a core sense of moral responsibility..." The

notion of moral responsibility Pereboom (2014, p. 2) discusses is the sort that “is set apart by the notion of *basic desert*...For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary.” Using characterizations such as these, Pereboom and other proponents of manipulation arguments have argued that causal determinism is incompatible with the sort of free will required for moral responsibility, where causal determinism is defined as the thesis that every event is determined by earlier events and the laws of nature. Though free will and moral responsibility are closely related, manipulation arguments have recently focused on moral responsibility. Therefore, while many of the concepts I discuss will also be applicable to the free will debate, I will focus solely on issues of moral responsibility for the remainder of this thesis.

Whereas Derk Pereboom and incompatibilists in general argue that determinism precludes moral responsibility, compatibilists argue this is not necessarily the case. Although according to compatibilism, determinism does not necessarily undermine moral responsibility, there certainly seems to be some factors which *do* undermine responsibility. For example, if an agent is coerced or physically forced to perform a certain act, then they would not be morally responsible. As a result of the somewhat complex position they hold, compatibilists have taken up the task of attempting to explain under what conditions an agent is morally responsible. Some popular compatibilist requirements for an agent to be considered morally responsible for an action are that an agent’s effective desire to act must appropriately conform with her second-order desires (Frankfurt 1971), the agent must be reasons-responsive (Fischer and Ravizza 1998), and the agent must be able to appropriately develop and revise over time the character traits that motivate her actions (Mele 1995; 2006; Haji 1998; 2009).

Given this understanding of the compatibilist view, there are two ways in which a manipulation argument might be problematic for compatibilism. One way a manipulation argument would be problematic is if it successfully described a scenario in which an agent lacks moral responsibility despite having met all compatibilist requirements for moral responsibility. This is typically what the first case of any manipulation argument attempts to show. In order for a manipulation case to be a successful counterexample to the compatibilist account of moral responsibility, the following conditions must be met.

Condition 1: All compatibilist requirements for moral responsibility must be met.

Condition 2: The relevant readers<sup>2</sup> of the manipulation case must intuitively find the manipulated agent either to lack moral responsibility.

Condition 3: Readers who have the intuition that the manipulated agent lacks moral responsibility or has diminished levels of moral responsibility must be properly conceiving all the relevant features of the manipulation case.

If any of these conditions are not met, then the manipulation case fails as a counterexample to compatibilism. For example, if all compatibilist conditions are not met (Condition 1), then the manipulation case could not succeed in showing how all of these compatibilist requirements are insufficient for moral responsibility, since the missing compatibilist condition might explain why the agent lacks responsibility. If readers of the manipulation case do not have the intuition that the manipulated agent lacks moral responsibility or at least has diminished moral responsibility for the action described (Condition 2), then the argument fails as an argument against compatibilism since it fails to provide the reader with solid grounds to suppose that there is a problem with compatibilism. Last, if the intuition that a manipulated agent lacks moral responsibility is the result of misunderstanding what role the compatibilist capacities play in the situation or

---

<sup>2</sup> I discuss which readers should be considered “relevant” in section 3 below.

misunderstanding some important aspect of the manipulation case (Condition 3), then this intuition would carry no weight in demonstrating that compatibilist requirements are insufficient to secure moral responsibility. In order to correctly provide a counterexample for a philosophical position, one must be objecting to the actual position, not to a straw man version or confused version of that position. If, however, these three conditions are met, then it seems we are justified in using intuitions about a manipulation case to show that the compatibilist account of moral responsibility fails to capture all that is required for an agent to be morally responsible.

It is important to note, though, that introducing a manipulation case which meets these three conditions merely provides a counterexample to the currently-proposed compatibilist account of moral responsibility. That is, this manipulation case alone does not undermine the possibility that determinism and moral responsibility are compatible. Rather, it merely demonstrates that the requirements which compatibilists *currently* cite as being necessary for moral responsibility are not sufficient for moral responsibility. In response to a successful manipulation argument which meets the three conditions I described above, a compatibilist can always respond by adding another requirement that is necessary for moral responsibility.

For example, suppose an agent, Bob, meets all compatibilist requirements for moral responsibility, but Bob also has a migraine that causes his reasoning to be slightly altered in such a way that he decides to kill David and he would have not made this decision if he had not had this migraine. If the first three conditions are met (readers correctly conceive of Bob as meeting *all* compatibilist requirements for moral responsibility and yet they intuit that Bob is not morally responsible), then we would have found a counterexample to the compatibilist account of moral responsibility. However, this does not support incompatibilism since this case says nothing about whether *determinism* is incompatible with moral responsibility. Rather it would only demonstrate

that *Bob's migraine* is incompatible with Bob being morally responsible. Since Bob's migraine has nothing to do with determinism, this case does not support incompatibilism. Therefore, the compatibilist could add that in order to be morally responsible, an agent must *not* have the kind of migraine Bob had. This would fix the complication for compatibilism while being irrelevant to discussions of incompatibilism. Another example of a manipulation case that meets these three conditions but does not support incompatibilism is Mele's (1995) case of an agent who has someone else's values implanted into them overnight. Mele seems to have succeeded in finding a problem with compatibilist accounts of moral responsibility since they had previously failed to consider that in order to be morally responsible, one must have a causal history that allows them to appropriately revise and develop their character over time. However, Mele does not necessarily undermine compatibilism completely or succeed in arguing for incompatibilism with this case.

Thus, in addition to demonstrating that there is a hole in the compatibilist account, proponents of manipulation cases who intend to argue for incompatibilism need to further argue that the responsibility-undermining feature present in the manipulation case is also present in a deterministic universe. If the proponent of a manipulation argument can successfully demonstrate that the manipulated agent's responsibility is undermined by some feature which is also necessarily present in a deterministic universe, then the entire compatibilist view of moral responsibility would be undermined since moral responsibility would not be compatible with determinism. Therefore, in addition to merely finding a complication for compatibilism, showing that compatibilism is untenable requires that manipulation arguments meet a fourth condition:

Condition 4: The responsibility-undermining feature of the manipulation case must be a feature of determinism.

Satisfying Condition 4 would result in a manipulation argument that not only provides a counterexample to specific proposed compatibilist accounts, but that also provides a successful

positive argument for incompatibilism. In order to demonstrate that Condition 4 is met, proponents of manipulation arguments attempt to show that manipulated agents are not relevantly dissimilar to agents in a deterministic universe. In order to demonstrate the similarity between a manipulation case and a world where everything is causally determined by the past and natural laws, proponents of manipulation cases often follow up the presentation of a manipulation case with similar cases or cases involving only determinism and argue that there are no relevant dissimilarities between the former and latter cases. Such manipulation cases can roughly be understood to have the formulation below (McKenna 2008; Mele 2008). I will refer to this formulation as MA:

- (P1) The manipulated agent is not morally responsible.
- (P2) There is no difference relevant for moral responsibility between the manipulated agent and an agent in a deterministic universe.
- (C) Therefore, an agent in a deterministic universe is not morally responsible.

Assuming that the manipulation case described in P1 meets the three conditions I spelled out above, the truth of P1 would demonstrate that the compatibilist requirements for moral responsibility are insufficient. Given P1, the truth of P2 would demonstrate that the responsibility-undermining feature present in a manipulation case is also present in a case involving only determinism and, thus, compatibilism is untenable. In order to defend P2, the responsibility-undermining feature in all cases must be a feature which is present in a case involving only determinism (ideally, determinism is the *only* feature present in the last case of a manipulation argument). If the only relevant feature in the last case of a manipulation argument is determinism, and there are supposed to be no morally relevant differences between cases, then all features of the first case which are not present in a deterministic universe must be irrelevant to moral responsibility. If it turns out that any features that are present in earlier cases are relevant to moral

responsibility and are not present in a case involving only determinism, then P2 does not hold and the manipulation argument is unsuccessful. Furthermore, while cases may share many insignificant properties that are irrelevant to determinism, it must be shown that these shared features are not driving intuitions. However, it is one of the goals of this thesis to impress upon the reader the vulnerability of intuitions – that intuitions are often significantly affected by unintended and seemingly irrelevant features of thought experiments. Therefore, these seemingly insignificant features are in fact highly significant if they significantly affect intuitions and moral judgments. Therefore, in order for P2 to hold, no features other than those present in a causally determined universe can influence the relevant readers’ intuitions and judgments of moral responsibility.

Another way to describe the general framework of manipulation arguments is to understand that they are accepting a couple of general principles in order to inductively conclude that determinism rules out moral responsibility. For instance, proponents of manipulation arguments are appealing to intuitions in order to determine what does and does not undermine moral responsibility. Therefore, at least *within the realm of moral reasoning* we can assume that they accept what I will call the Intuition Principle:

(IP) If people consistently intuit that X, then there is good (though defeasible) reason to believe that X is true.

I am not assuming that this reasoning applies to all intuitions, but this seems to be a common methodology for justifying claims about morality and moral responsibility and is employed by proponents of manipulation cases as well as many other philosophers debating questions about morality.

Secondly, though some have argued that manipulation arguments do not need to offer an explanation for the intuitions in response to the cases (Mele, 2005; 2008), I argue, in agreement with Mickelson (2015), that providing an explanation for what drives intuitions of non-

responsibility is of utmost importance. Those who use manipulation arguments to support incompatibilism need to demonstrate that some feature of *determinism* rules out the possibility of an agent's being morally responsible. In order to show which feature undermines responsibility, one needs to demonstrate that the manipulated agent's lack of moral responsibility is explained by the presence of some deterministic feature. Thus, in order to argue that determinism precludes moral responsibility, proponents of manipulation cases should also endorse what I consider the Incompatibility Principle:

(IncP) If A best explains why an agent is not morally responsible, then we have good reason to believe that A is incompatible with moral responsibility.

If we fill in X and A in these two principles with features of manipulation cases by adding that people reliably intuit that manipulated agents are not morally responsible and that features of determinism best explain why the manipulated agents lack moral responsibility, then the manipulation argument can be roughly formulated as below:

- M1) If people consistently intuit that the manipulated agent is not morally responsible, then there is good reason to believe the manipulated agent is not morally responsible.
- M2) People consistently intuit that the manipulated agent is not morally responsible.
- M3) Therefore, there is good reason to believe the manipulated agent is not morally responsible.
- M4) If features of determinism best explain why the manipulated agent is not morally responsible, then we have good reason to believe features of determinism are incompatible with moral responsibility.
- M5) Features of determinism best explain why the manipulated agent is not morally responsible.
- C) Therefore, we have good reason to believe features of determinism are incompatible with moral responsibility.

This presentation of manipulation arguments make explicit some of the inductive reasoning that is required in order for the manipulation argument to succeed. The above formulation



highlights two things: (1) our reliance on intuitions to infer conclusions about moral responsibility, and (2) just how much hinges on precisely what best explains why manipulated agents lack moral responsibility. Since what motivates intuitions about manipulation cases is extremely important for manipulation arguments, these arguments, and ones with similar reliance on intuitions, must take intuition-affecting psychological influences into consideration in order to be immune from the objection that we only find them compelling because of those psychological influences. I argue that ignoring these unconscious psychological influences and heuristics can undermine one's argument.

In order to provide a specific example of an argument that is threatened by neglecting to account for such factors, I present Derk Pereboom's four-case manipulation argument. Using the presentation of manipulation arguments I have described above, this thesis will focus on rejecting premise M5 by arguing that features of determinism do *not* best explain why the manipulated agent lacks moral responsibility. Rather, I argue that something independent of the features of determinism best explains why people judge that manipulated agents lack moral responsibility. Therefore, something other than determinism would be incompatible with moral responsibility and the manipulation argument for incompatibilism is unsuccessful. Given the way in which Derk Pereboom's manipulation argument is presented, it seems extremely likely that seemingly irrelevant psychological influences, such as the order in which he presents his cases, provide a better explanation than the one which Pereboom offers for why readers intuit that determined agents are not morally responsible.

### 3 PEREBOOM'S FOUR-CASE MANIPULATION ARGUMENT

To demonstrate that once determinism is properly understood, compatibilism fails, Derk Pereboom (2014) presents a manipulation argument. Pereboom attempts to convince his audience that even in cases when all compatibilist requirements for moral responsibility are met, agents can still lack moral responsibility. He also intends to convince readers of his argument that moral responsibility is undermined by features of causal determinism. While it is important to know exactly who Pereboom's intended audience is and to what degree such an audience actually exists, determining these matters is difficult and I will not spend much time speculating. However, Pereboom seems to be targeting both compatibilists and philosophers who are agnostic about the free will debate. He states, "...the manipulation argument aims to persuade the natural compatibilist and the agnostic their resistance to incompatibilism is best given up." (2014, p. 81) Therefore, I will assume for the remainder of this thesis that Pereboom intends the readers of his argument to be either compatibilists or agnostic philosophers.

In order to convert these natural compatibilists and agnostic readers to incompatibilism, Pereboom presents four cases. Each case involves an agent, Plum, who is causally determined by factors beyond his control to kill another agent, White. Additionally, in each case Plum satisfies all purported compatibilist requirements for moral responsibility.<sup>3</sup> Case 1 reads:

A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-

---

<sup>3</sup> Pereboom asserts that in all four cases Plum satisfies the requirements which Hume (1739/1978), Harry Frankfurt (1971), John Fischer and Mark Ravizza (1998), Jay Wallace (1994), and Alfred Mele (1995; 2006) have argued are necessary for an agent to be considered morally responsible.

order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort. (Pereboom 2014 p. 76-77)

Case 2 is similar to Case 1 “except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic,” (2014, p. 77) as opposed to Plum's being manipulated just before he reasoned about his situation as occurs in Case 1. Again, Plum maintains all compatibilist requirements for moral responsibility and yet Plum is intuitively not morally responsible for his decision to kill White. In Case 3, it is the training practices of Plum's community, which were completed before he developed the ability to prevent or alter these practices, that causally determined the nature of his deliberative reasoning process such that he reasons egoistically and kills White (2014, p. 78). In Case 4 of Pereboom's manipulation argument, Plum is an ordinary human being, raised in normal circumstances in a world where everything, including Plum's egoistic decision to kill White, is causally determined by its past states and the laws of nature. Again, in all four cases Plum satisfies *all* purported compatibilist requirements for moral responsibility *and* Plum's actions are ultimately determined by factors outside of his control.<sup>4</sup>

Given this presentation, whether Pereboom's argument successfully poses a problem for compatibilist accounts of moral responsibility depends on its meeting the three conditions I

---

<sup>4</sup> While Demetriou (2010) argues it may be metaphysically impossible for the manipulation Pereboom describes to occur without inviting either a hard- or soft-line response, for the purposes of this thesis, I will assume a metaphysically coherent interpretation of Pereboom's manipulation does exist.

presented earlier; all compatibilist requirements for moral responsibility must be met, readers must intuitively find Plum to lack moral responsibility, and readers who have the intuition that Plum lacks moral responsibility must be properly conceiving all of the features of these cases. Pereboom would argue that all of his four cases meet these requirements, and thus, present a serious problem for compatibilism.

Nonetheless, it seems Pereboom is attempting to do more than provide a counterexample for compatibilism. We can see that Pereboom additionally intends to make a positive argument for incompatibilism by demonstrating that it is a feature of causal determinism that is incompatible with moral responsibility.

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, p. 79)

By attempting to show that a feature of determinism (Plum's actions are ultimately causally determined by factors beyond his control) explains intuitions of Plum's non-responsibility across all four cases, Pereboom is attempting to demonstrate that his argument also meets Condition 4, namely, that the responsibility-undermining feature of the manipulation case is a feature *of determinism*.

Before I present my argument, it is worth noting that there are in fact reasons to doubt whether the first three conditions, which are necessary in order for any manipulation case to succeed in poking holes in the compatibilist account of moral responsibility, are actually satisfied. For instance, some have argued that all of the compatibilist requirements for moral responsibility either cannot be met or are not met in manipulation cases like the one Pereboom describes (Waller 2013; Demetriou 2010). If this is the case, then Condition 1 is not satisfied. Also, there is some

evidence that readers don't actually have the intuition that Plum lacks moral responsibility (Feltz 2013). If this is the case, then Condition 2 is not satisfied. Lastly, experimental philosophy has provided reason to believe that readers are easily confused about what determinism properly entails (Murray and Nahmias 2014) and that most readers fail to understand manipulated agents as having all of the necessary compatibilist requirements for moral responsibility (Sripada 2011). If this is the case, then Condition 3 is not satisfied and hence these intuitions cannot be used by Pereboom to support his anti-compatibilist argument. However, while these are significant problems for Pereboom's argument, I wish to draw attention to the problem that arises from potentially failing to meet Condition 4. Specifically, I argue that the intuition that Plum is not morally responsible in all four cases is not motivated by some feature of determinism, but by unconscious psychological influences that are not relevant to the truth of determinism (like order effects). If I am right, then Pereboom's four cases fail to meet Condition 4 and, thus, his argument for incompatibilism fails.

Given that Pereboom is attempting to show both that the compatibilist conditions for moral responsibility are insufficient and that determinism is incompatible with moral responsibility, a single feature of these cases (Plum's actions are ultimately determined by factors outside his control) needs to explain why it is that individuals intuitively find Plum not morally responsible. If intuitions that Plum lacks responsibility result from any other aspects of the argument, then Pereboom's argument fails. This is because something independent of the features of determinism would best explain why people judge that Plum lacks moral responsibility, and thus, something other than determinism would be incompatible with moral responsibility.

In all four of Pereboom's cases Plum's actions are certainly determined by factors outside his control. However, just because all four cases share this property, this does not necessarily

mean that this is the feature is what motivates the intuition that Plum is not morally responsible. As Mele (year, p. 79) cleverly explains, we should find Pereboom's explanation about his four cases,

...no more plausible than the claim that the best explanation for Scarlet's car's being damaged in the following three cases is that it was struck by an object that was, among other things, *wet*: (Case 1) Scarlet's car was struck by a falling large wet lead pipe and was damaged as a result; (Case 2) Scarlet's car was struck by a falling large wet wrench and was damaged as a result; (Case 3) Scarlet's car was struck by a falling large wet metal candlestick and was damaged as a result. (In each case, the object fell ten metres.) The claim that I have just invited you to recall immediately precedes Pereboom's assertion that 'Because Plum is also causally determined in this way in Case 4,...we should conclude that here too Plum is not morally responsible for the same reason' (116). As it happens, in case 4 of the Scarlet chronicles, her car was struck by a falling large wet sponge. Peacocke concludes that Scarlet's car is damaged in this case too. But, of course, she is wrong. It was such things as the hardness and weight of the falling objects, not their wetness, that did the work.

Mele's point here is that the mere fact that Plum's actions are ultimately causally determined by factors outside his control in the first three cases and the fact that this feature is shared in the fourth case does not entail that we should expect the shared feature to do work in motivating intuitions regarding Case 4. It is possible that there are other features of the first three cases that do work in undermining responsibility that do not necessarily apply to a case involving only determinism.

It is essential for the success of Pereboom's argument for incompatibilism that his explanation for what motivates these intuitions is actually what motivates intuitions, and Pereboom must demonstrate that some feature of determinism is what actually does the work in his four-case argument. In the next section, I argue that Pereboom's presentation of the four-case argument leads to certain, largely unconscious, psychological influences driving our intuitions about Plum's not being morally responsible. Since the effects of these psychological influences lead to order effects, I argue that order effects can provide a better explanation for why readers have the intuition

that Plum is not morally responsible in Case 4 and expose a weakness in Pereboom's argument which potentially undermines his argument.

Before moving forward and arguing for an alternative explanation for what drives our intuitions in Pereboom's cases, I would first like to point out a number of potential confusions and offer a disclaimer about what I am *not* attempting to do. I am not putting forth a positive argument for compatibilism, nor am I even necessarily defending any compatibilist view. I am merely highlighting problems with Pereboom's argument for incompatibilism. Also, I am not arguing that because some psychological factors influence intuitions and moral judgments, philosophers should refrain from ever appealing to intuitions and thought experiments. I am merely providing evidence to suggest that *if* philosophers are going to rely on intuitions about thought experiments, then they will need to precisely determine what drives these intuitions if they wish to use those intuitions as premises in their arguments. Philosophers who rely on intuitions need to take unconscious psychological influences into account when explaining what drives intuitions.

Furthermore, I am not offering a hard-line response to manipulation arguments and arguing that Plum actually *should* be considered morally responsible in all four cases (McKenna 2008), nor am I necessarily taking a soft-line response and arguing that there is a relevant dissimilarity between two of the cases which allows us to consider Plum not morally responsible in Case 1 but morally responsible in Case 4 (Demetriou 2011; Waller 2013).<sup>5</sup> While one might think that I need to either take a hard-line response and argue that P1 of MA is false or take the soft-line response and argue that P2 of MA is false, these positions are only required if one is attempting to argue

---

<sup>5</sup> While my argument allows for one to reject P2 and therefore, in some sense, provides a soft-line response to Pereboom's argument, I do not intend for my argument to *only* be a soft-line response. Rather, this is a plausible but not central product of this thesis. I focus my attention on arguing that Pereboom fails to meet Condition 4 and does not provide the best explanation for intuitions regarding his four cases.

that there is no problem with the compatibilist account of moral responsibility. However, I am not necessarily arguing that the manipulation arguments do not elucidate problems for compatibilism. Rather, I am showing Pereboom's manipulation argument fails as an argument *for incompatibilism*. That is, though Pereboom's four-case argument may or may not demonstrate a complication for the compatibilist account, I argue this complication would not succeed in demonstrating the truth of incompatibilism. In arguing this, I raise methodological concerns like those Mele (2005) addresses and call into question Pereboom's explanation for why we find Plum intuitively lacking moral responsibility. I aim to undermine Pereboom's argument for incompatibilism by offering a better explanation for these intuitions.

#### **4 ORDER EFFECTS AS AN ALTERNATIVE EXPLANATION**

In this section, I argue that order effects provide a better explanation for what motivates judgments of Plum's non-responsibility in the four-case manipulation argument than Pereboom's explanation. That is, I argue that the intuitions readers have about Pereboom's four cases are affected by the order in which the cases are presented, and that these intuitions would be different if they were presented in a different order, or if they were presented independently of one another. After providing evidence that the order in which Pereboom's four cases are presented affects judgments about the extent to which Plum is morally responsible, I will discuss specific features and psychological mechanisms that likely lead to these order effects. Though knowledge of the fact that human psychology is often subject to order effects may be reason enough to suspect that such a phenomenon is at play in Pereboom's four-case argument, I will also discuss the possibilities that an agency-detection mechanism, the presence of agential intent in earlier cases, or emotional responses to Case 1 are motivating order effects, as there is empirical evidence which



suggests these psychological influences unconsciously motivate intuitions and result in order effects.

Alex Weigmann, Yasmina Okan, and Jonas Nagel (2012) have demonstrated that the order in which trolley dilemmas are presented significantly influences judgments of moral permissibility.<sup>6</sup> After presenting participants with five variations of the trolley dilemma, which differed only in what the life-saving action was, subjects' responses were affected by the order in which the cases were presented.<sup>7</sup> Weigmann et al. concluded, “judgments would be most likely transferred if the initial rating was strongly negative” (2012, p. 825). That is, when readers had a strongly negative judgment towards the first case, this judgment was likely to affect judgments of later cases. This highly negative first case resulted in consistently more negative judgments of moral permissibility (relative to judgments of these cases presented on their own). Though I will provide reasons for *why* readers might have strongly negative reaction to Pereboom’s Case 1, it is enough here to note that readers do have negative reactions to Case 1 and do judge Plum to lack moral responsibility (Feltz 2013). Given these negative reactions to Case 1, I argue that reading Case 1 first affects judgments of moral responsibility of later cases much in the same way Wigmann et al. observed order affected judgments towards trolley dilemmas. Therefore, if the Cases were presented in a different order – for instance, in reverse order – judgments would be significantly altered.

---

<sup>6</sup>Trolley dilemmas are scenarios where a trolley train is out of control and on track to run over multiple workers. However, someone has the option of choosing to sacrifice the life of one person to save the multitude, and these scenarios vary according to how that sacrifice must be carried out.

<sup>7</sup> The potentially life-saving actions were: pressing a switch that will redirect the train that is out of control to a parallel track where one person will be run over; redirecting an empty train that is on a parallel track onto the main track to stop the train, running over a person that is on the connecting track; redirecting a train with a person inside that is on a parallel track onto the main track to stop the train; pushing a button that will open a trap door that will let a large person on top of a bridge fall and stop the train; push the large person from the bridge to stop the train.

While one might assume the experienced agnostic philosopher would not be affected by the order in which cases are presented, Schwitzgebel and Cushman (2012) found that order of presentation influenced the moral judgments of philosophers *more* than it did non-philosophers! Furthermore, this effect persists among philosophers who specialize in ethics (Schwitzgebel and Cushman forthcoming). Not only do Schwitzgebel and Cushman's findings suggest that philosophers need to take the salience of order effects seriously, but they could possibly provide reason to be *more* worried about these effects occurring in philosophy than in other areas.

If it turned out that order effects better explain why we find Plum not morally responsible in later cases, then Pereboom would fail to provide the best explanation for these intuitions which are critical to his argument. I am not claiming that, in principle, order effects undermine philosophical arguments. However, if the argument is one that depends on being able to correctly identify what motivates intuitions about certain cases, as Pereboom's argument for incompatibilism is, then the argument fails if it incorrectly identifies what motivates intuitions. Remember that the success of Pereboom's argument for incompatibilism hinges on satisfying Condition 4 and showing that some feature of determinism motivates intuitions of non-responsibility. If I am right that these intuitions are not the result of some feature of determinism but of unconscious psychological influences that pick up on other largely irrelevant features of those four cases, then we lack reason to think that causal determinism is incompatible with moral responsibility and Pereboom's argument is unsuccessful.

#### **4.1 Agency-Detection Mechanism**

The first psychological mechanism that I argue influences judgments regarding Pereboom's four cases is an agency-detection mechanism. Given that Case 1 clearly describes agents (neuroscientists) acting upon Plum, I argue that an agency-detection mechanism becomes active while reading Case 1 and this causes readers to erroneously think in agential terms when reading

Case 4 and thinking about what determinism entails. This agency-detection mechanism would not be active if Pereboom presented Case 4 first, since only non-agential causal determinism acts on Plum in Case 4. Because such a mechanism would influence judgments of Case 4 when Case 4 is presented last, but not influence judgments when Case 4 is presented first, the order of presentation affects intuitions, and, thus, on my account this mechanism contributes to the effect that the order of presentation has on intuitions about these cases.

Scott Atran (2006) argues that human evolution has naturally selected for an innate and overly sensitive mechanism for detecting agents and agential properties. While this mechanism often beneficially and accurately identifies agents, Atran argues that it also causes humans to wrongly attribute agential properties to nearly any complex or uncertain situation or design. For example, Atran believes this overly sensitive mechanism explains why people often see faces in the clouds and are quick to believe in supernatural beings. The reason Atran believes our agency-detection mechanisms would be *overly* sensitive is because, evolutionarily speaking, it is much more costly to fail to attribute agency to a (potentially dangerous) agent than it would be to attribute agential features to non-agential things like clouds. That is, we are likely to identify agential properties in situations where there are none present because it has been evolutionarily beneficial for us to do so.

Such an agency-detection mechanism would become active when reading about the neuroscientists in Case 1 who determine Plum's actions. This would lead us to correctly attribute agential properties to the causal determinants of Plum's actions. However, given both the similarities between cases and Pereboom's continual insistence that there are no relevant differences between cases, this agency-detection mechanism would likely remain active in later cases, including Case 4 when Pereboom eventually removes agents and agential properties from

the picture entirely and replaces them with the complex structure of causal determinism. If this mechanism remained active, then readers would (likely unconsciously) attribute agential properties to the causal determinants of Plum's actions in Case 4. However, determinism has no such agential properties, and so if Pereboom's presentation of the cases causes readers to unwittingly assume that determinism has some form of agency, then Pereboom's presentation confuses the reader about the nature of determinism.<sup>8</sup>

If this overly sensitive agency-detection mechanism does in fact influence intuitions about Case 4, then the order in which Pereboom presents these cases has an effect on judgments of Plum's non-responsibility. Furthermore, this alternative explanation for intuitions would undermine Pereboom's goal of getting readers to properly understand the causal nature of determinism and result in the manipulation case failing to meet Condition 3. Since determinism, and therefore Case 4, does not involve agents or agential properties which influence Plum, it would be misguided for intuitions about Case 4 to be influenced by agency. If intuitions about Plum in Case 4 are motivated by an agency-detection mechanism responding to agency in earlier cases, as I argue they are, then these intuitions are unreliable and cannot be used to motivate Pereboom's argument.

#### 4.2 *Intent*

While the mere presence of agents in Case 1 might cause readers to judge Plum not morally responsible in Case 4, the intent of these agents also appears to contribute to the order effects. Phillips and Shaw (forthcoming) investigated how third-party intent (the intent of agents who causally determine how another agent acts but nonetheless are not necessarily acting or being affected by the action themselves) influences judgments of moral responsibility. First, they found

---

<sup>8</sup> In an unpublished manuscript, Neil Levy makes a similar argument, claiming that Pereboom's four-case manipulation argument only succeeds insofar as it activates an agency-detection mechanism which causes the readers to see determinism in agential terms.

that the presence of third-party intent does reduce judgments of blame.<sup>9</sup> Second, third-party intent *only* influenced judgments when the agent's actions match the intended action. Third, their results suggest that intent affects judgments of moral responsibility by altering the reader's causal perception. If Pereboom's four-case argument successfully alters one's causal perception only because third-party intent is present in earlier cases, then judgments of earlier cases are influencing judgments of later cases. If later judgments are being affected by earlier judgments, then the order of presentation again has an effect on our intuitions of non-responsibility. If intuitions of Plum's non-responsibility are the result of order effects, then we have an alternative explanation for these intuitions that is deeply problematic for Pereboom's argument.

To understand why it is problematic for third-party intent to alter our judgments here, consider the following: According to Pereboom, many people don't see determinism as ruling out the possibility of moral responsibility because they misunderstand the true nature of determinism. To remedy these misconceptions, "the manipulation cases are formulated so as to correct for inadequacy in the extent to which we take into account hidden deterministic causes in our intuitions about ordinary cases" (2014, p. 95). That is, manipulation cases are intended to expose to us the true causal nature of determinism and they attempt to alter how one perceives the causal implications of determinism. Phillips and Shaw hypothesize that manipulation cases can succeed in altering one's causal perception *only* when third-party intent is present and matches the action performed. Therefore, according to Phillips and Shaw's assessment, if a change in causal perception occurs, it is because readers understand there to be third-party intent present which matched the action. While Pereboom is clearly attempting to change the reader's causal perception, it would be mistaken to alter perceptions by getting readers to understand determinism

---

<sup>9</sup> These findings are consistent with Robyn Waller's (2013) argument that intent is a relevant difference between cases and affects judgments of moral responsibility.

as having any intent (or, for that matter, any other agential properties) since compatibilists and incompatibilists agree that this is the wrong way to conceive of determinism. This suggests that Pereboom's four cases elicit the desired intuitions by confusing readers about the true nature of determinism. Again, since Pereboom needs the responsibility-undermining feature of all of his four cases to be a feature of determinism, and intent is not a feature of determinism, if intent is influencing judgments of non-responsibility, then Pereboom's argument for incompatibilism faces a serious problem.

While the concern outlined above is certainly problematic for Pereboom's argument, it is worth noting that in order for my argument to succeed, intent doesn't necessarily need to confuse readers about the true nature of determinism. Rather, I merely need to demonstrate that the intent, along with other unconscious psychological influences, leads to order effects that influence judgments, and that these order effects explain intuitions of non-responsibility better than the mere fact that Plum's actions are ultimately determined by factors over which he has no control.<sup>10</sup>

### 4.3 *Emotional Responses*

Another psychological influence that likely causes the order of presentation to have an effect on judgments is emotional engagement with features present in Case 1. The first case of the four-case argument involves agential intent, an abnormal bodily violation (brain manipulation), and an abnormal social violation (manipulation). Reading vignettes that contain intent, abnormal bodily violations, and abnormal social violations has been shown to elicit emotional responses of disgust and anger (Giner-Sorolla 2011; Haidt 2003). Also, engaging emotionally with such vignettes has been shown both to be correlated with particular moral judgments (Greene 2001), as well as to

---

<sup>10</sup> In a response to Mele's criticisms, Pereboom (2014, 82) argues that even if these intentional agents "were replaced by force fields or machines that randomly form in space that have the same deterministic effect on Plum as the manipulators do, the intuition that Plum is not morally responsible persists." While I remain skeptical of this claim, it is interesting that Pereboom chooses not to make this replacement and he only mentions such a possibility after priming the reader with cases involving intentional agents.

influence moral judgments (Haidt 2003; Guiseppe et al. 2012), even when these emotions are primed non-consciously and automatically (Valdesolo and DeSteno 2006).<sup>11</sup> Furthermore, responding to a vignette with anger or disgust has been shown to affect judgments and behavior for some time after reading the vignette (Plaisier and Konijn 2013; He et al. 2013).

In light of such evidence, it is likely that readers of Pereboom's four-case argument would have a strongly negative emotional response to Case 1, and that this highly negative response would subsequently keep influencing our judgments all the way through to Case 4. Insofar as one's emotions are negatively responding to agential intent, body violations, or social violations, and not to the fact that Plum's actions are causally determined by factors over which he has no control, emotional engagement serves as a plausible confounding variable for what explains our judgments. That is, if our intuitions about Plum are the result of responding to emotional-priming factors that are irrelevant to determinism, then it isn't a feature of determinism that drives moral judgments, as Pereboom argues. Since features of Case 1 elicit emotional reactions, emotional engagement with features present in Case 1 influence judgments of later cases, thus leading to order effects taking place. These order effects, again, serve as an alternative explanation for intuitions of Plum's non-responsibility in Case 4 and thereby threaten the success of Pereboom's four-case manipulation argument. In summary, I argue that given Pereboom's presentation of his four-case manipulation argument, features only present in earlier cases (agents, third-party intent, abnormal body and social violations) are triggering certain unconscious psychological mechanisms that drive judgments of Case 4, thus resulting in order effects. Since these features are independent of any features of determinism, it would seem that something other than

---

<sup>11</sup> Haidt (2001) argues that in most circumstances, emotional engagement is the primary cause of moral reasoning. While this may or may not be the case, for my argument to work, it only needs to be the case that emotional engagement influences judgments of Pereboom's four cases.

determinism is motivating our moral intuitions and, therefore, Pereboom's argument for incompatibilism fails.

There may be additional psychological influences that drive order effects which I have not discussed. For example, intuitions might also be swayed by one's own demands for consistency across cases. That is, if readers are torn between whether we should consider Plum morally responsible in later cases and understand that the cases are similar in many ways, – and again, note that Pereboom reminds the reader that the cases *are* alike in all other regards – they are likely to answer consistently with how they did in earlier cases. Thus, if readers commit to saying that Plum is not morally responsible in Case 1, it is likely that they would answer consistently across cases merely for the sake of having consistent responses.

Another possible factor that might influence judgments, which Pereboom does not address, is that readers might have intuitions of non-responsibility about the four cases simply because Pereboom makes suggestions about what intuitions readers ought to have. If, immediately after someone reads one of Pereboom's four cases, this person has not yet decided whether or not to judge Plum morally responsible, their judgment could be significantly altered by what immediately follows the description of Cases 1 and 2; Pereboom (2014, p. 77) claiming that "...intuitively, he is not morally responsible for his decision." Furthermore, insofar as readers are unsure whether to judge Plum as morally responsible or not, it seems likely that readers are swayed by Pereboom because he is understood to be some kind of authority figure on what one ought to think about these cases. Readers might agree with Pereboom's judgment because he created these four cases and should therefore know enough about the relevant features and arguments to understand what the correct intuition is in a way that the reader themselves might think they are less capable.



If *any* such influences, either collectively or on their own, better explain why we (or the “agnostic reader”) find Plum intuitively not morally responsible in case 4, then Pereboom’s argument is unsuccessful because what precludes moral responsibility is not exclusively Plum’s actions being ultimately causally determined by factors outside his control (Condition 4 of Pereboom’s manipulation argument would not be satisfied). Therefore, Pereboom, like anyone else attempting to make claims about what drives intuitions, needs to take unconscious psychological influences seriously because, as I have now demonstrated, neglecting to acknowledge seemingly irrelevant influences, such as order effects, can undermine one’s argument.

In order to further elucidate importance of taking order effects into account, consider also the order in which information is presented *within* each case. Those who use manipulation arguments to challenge compatibilism (Mele 1995; Pereboom 2014) begin the description of each case by explaining the manipulation and only afterwards mention how the manipulated agent still satisfied the compatibilist requirements for moral responsibility. Contrastingly, those who draw compatibilist conclusions from manipulation cases (Sripada 2011) begin the description of their cases by explaining how the agent meets all compatibilist capacities and later mention the manipulation that takes place. The fact that the order of presentation within cases seems to change in accordance with the conclusions one ends up drawing is telling, and it nicely complements my claim that the order of presentation likely has a significant effect on our intuitions. Though I am not suggesting that any of the proponents of these manipulation cases knowingly present their cases this way to manipulate the readers, I share Wienberg’s concern about such order effects, and I too worry “that philosophers might be manipulating their own results without even being aware that such manipulation is taking place” (Weinberg et al. 2008).

## 5 OBJECTIONS AND EVIDENCE

Above I argued that in his four-case manipulation argument, Pereboom fails to take into account several salient and largely unconscious psychological influences that have been shown to affect intuitions. I argue that his argument does not elicit judgments about moral responsibility in the right way for those intuitions to support an argument for incompatibilism. More specifically, I have argued that intuitions of Plum's non-responsibility are best explained by factors other than the recognition that Plum's actions are causally determined by factors outside of his control. Rather, a better explanation involves the presence of order effects that are driven by certain psychological influences which readers are largely unaware of, such as an agency-detection mechanism, third-party intent, and negative emotional engagement with earlier cases. In this section I will consider objections to the argument I presented above.

### 5.1 *Order Effects Are Intended*

First, one might be tempted to object to my argument by saying something like the following: "Of course order effects sway intuitions in Pereboom's favor. The whole point of the four-case argument is to show people that when the features of determinism are presented less abstractly, we see that determinism undermines moral responsibility in the same way manipulation does. Therefore, the emotional responses and initial judgments about Case 1 *should* transfer over and influence intuitions about Case 4 because they allow us to think of these cases in the same way and with the same types of attitudes."

In response to this objection, I would first point out that insofar as Pereboom's four-case manipulation argument is to be understood as one that requires providing the best explanation for intuitions, the argument only works if Pereboom's explanation is actually *the best*. Therefore, if the fact that Plum's actions are ultimately being determined by factors outside his control is what

drives intuitions, certainly order effects are not problematic for Pereboom. If anything else motivates intuitions, then the argument simply doesn't work.<sup>12</sup> Therefore, it is important to know which feature of these cases do and do not motivate intuitions of non-responsibility.

Mele (2005; 2008) has argued that readers would judge Plum not morally responsible even if the causation in these cases was indeterministic, and that demonstrating this this would show that determinism is not what motivates intuitions about the four cases. If Mele is right and deterministic causation isn't what drives intuitions, then these judgments must be sensitive to other factors. I presented a few likely candidates for which features of these cases influence intuitions regarding Case 1: the presence of agents, third-party intent, and emotionally responding to manipulation. I also provided reason to believe that if the factors I discuss are what motivate intuitions about Case 1, then it's highly likely that order effects will take place and intuitions of non-responsibility will remain consistent across cases. Whereas Pereboom relies on his explanation of the cases to provide evidence that his explanation is best, I have provided a significant amount of empirical research which suggests an alternative explanation would better explain the our intuitions about the four cases. Therefore, while Pereboom might have intended the order in which the presentations are introduced to affect intuitions, this would be a mistake unless it somehow allowed the reader to better understand determinism, which I provide reason to believe is not the case.

As a second response to the objection that order effects should occur, I'd point out that if order effects are intended to take place and we are supposed to understand Case 1 and Case 4 in roughly the same way, then Pereboom is likely confusing the reader about the true causal nature

---

<sup>12</sup> For example, in all four cases, the agent being discussed is named Plum. If (for some strange reason) readers will always judge that Plum is not morally responsible for the sole reason that his name is Plum, then every other feature of these cases is irrelevant and says nothing about what rules out moral responsibility.

of determinism. As discussed earlier, if the intuition that Plum is not morally responsible in Case 4 is residually influenced by the presence of agents or third-party intent in Case 1, then the intuitions about Case 4 are misguided, since determinism has no agential properties or intentions. Furthermore, concepts like determinism and moral responsibility are complex but do not seem to be so confusing that the “agnostic philosopher” can only grasp them through an analogous thought experiment. Therefore, it seems that while Pereboom’s use of the four-case argument is superfluous to begin with, by incorporating many additional concepts (like intentional and agential manipulation) Pereboom is only confusing readers more about what determinism entails. Readers of Pereboom’s four-case argument are likely conflating features such as agency and intent with determinism in Case 4 and, thus, are confused about the true nature of determinism. However, even if this is not the case, it seems extremely unlikely that judgments are best explained by the single feature Pereboom addresses, given the many other features present in Case 1 that are known to engage psychological mechanisms that lead to order effects and alter judgments of later cases.

### 5.2 *Explaining Intuitions is Unimportant*

Secondly, one might argue that by presenting Pereboom’s four-case argument as an argument that requires an explanation for what motivates intuitions, I am misrepresenting it. After all, Pereboom’s four-case argument is a manipulation argument and some have argued that manipulation arguments can operate without any explicit premise that attempts to explain what motivates the non-responsibility intuition (Mele 2008; Todd 2012). Many others have assumed that Pereboom’s argument ought to be conceived of in this way as well (Feltz 2013; McKenna 2008; Mele 2005), and there seems to be some evidence that this is what Pereboom intends. For example, in Pereboom’s most recent presentation of his four-case argument, he argues,

It’s highly intuitive that Plum is not morally responsible in Case 1, and there are no differences between Cases 1 and 2, 2 and 3, and 3 and 4 that can explain in a principled way why he would not be responsible in the former

of each pair but would be in the latter. We are thus driven to the conclusion that he is not responsible in Case 4. (2014, p. 79)

This passage might lead one to assume Pereboom's argument is similar to other manipulation arguments which roughly follow the formulation I described as MA. MA does not include any premises about what explains judgments of non-responsibility and so one might think such an explanation is superfluous. However, in addition to running counter to Pereboom's stated intentions, I argue, as Mickelson (2015) does, that understanding a manipulation argument such as Pereboom's to operate without any explanation for what drives intuitions about manipulation cases would be problematic.

In order to understand why an explanation for intuitions is needed for the success of a manipulation argument, it is helpful to further investigate the methodology being employed with a manipulation argument. The goal of manipulation arguments is to derive truths about moral responsibility from intuitions about moral responsibility. That is, proponents of manipulation cases present the reader with scenarios, and (assuming the three conditions I described earlier are met) the proponent concludes that agents are not morally responsible. If everyone gets the same intuition and the intuition is well-founded, then we would accept the conclusion that moral responsibility really is incompatible with determinism. The point of manipulation cases then, is to find a way to demonstrate that moral responsibility is *intuitively* incompatible with determinism; by deriving conclusions about moral responsibility from intuitions about moral responsibility. Therefore, I argue that proponents of manipulation arguments endorse the Intuition Principle (If people consistently intuit X, then there is good reason to believe X is true.) Given this methodology, what drives intuitions about moral responsibility is extremely relevant to moral responsibility. Therefore, if our intuitions about manipulation cases are influenced by features of the cases that are independent of determinism and we accept the general methodology of

manipulation cases, then these features independent of determinism become relevant to moral responsibility in virtue of their ability to affect our intuitions.

Additionally, in order for a manipulation argument to succeed, it must be the case that there are no relevant differences between cases. That is, the feature which undermines responsibility in the first, second, or third case, must be the same feature which undermines responsibility in a case involving only determinism. However, in order to determine which features of a manipulation case are relevant for moral responsibility, one must explain what in fact influences intuitions about these cases. If intuitions about Cases 1, 2, or 3 are influenced by factors that are not present in Case 4 (agents, intent, or emotionally responding to agents intentionally manipulating Plum) and if we understand features that influence intuitions to be relevant for moral responsibility, then these features become relevant. If such features provide a relevant difference between cases, then P2 of MA (i.e. that there is no difference between the manipulated agent and an agent in a deterministic universe that is relevant for moral responsibility) does not hold. For example, if the best explanation is the intuition that Plum is not morally responsible in Case 1 is that the intentions of the neuroscientists preclude Plum's moral responsibility, then there is a relevant difference between Case 1 and Case 4. If the best explanation for the intuition of non-responsibility is that there are agents present who determine Plum's actions in Case 1, then M5 does not hold and again, there is a relevant difference between Case 1 and Case 4.

It seems likely that if one were presented with a case in which an agent were successfully and intentionally manipulated by agents to do something, regardless of whether determinism were true, we would likely still have the intuition that Plum is not morally responsible. If we suppose that what influences judgments of moral responsibility is relevant for moral responsibility, then one might argue that determinism is not incompatible with moral responsibility. Rather, some

other factors that are present in the four-case argument would be incompatible with moral responsibility.

Given that proponents of manipulation arguments are using intuitions regarding moral responsibility to draw conclusions about responsibility, we can use this methodology to determine whether or not features of determinism are actually what drive intuitions of non-responsibility. If features of determinism aren't what leads to the intuition that Plum lacks moral responsibility then it doesn't seem determinism is what rules out moral responsibility in these cases and, the argument for incompatibilism fails. As with most thought experiments, Pereboom's four cases involve many features. While all but a select few features are intended to be relevant, it is always possible that some other features which were not intended to affect intuitions do in fact motivate readers to judge that Plum is not morally responsible. It is thus crucial to be able to sort out which features are important from those which are not. Although there may be some feature present in all four cases – for instance, the protagonist's name is "Plum" in all four cases – this is not yet enough to determine whether this feature is what drives intuitions. While it may be the case that in all four of Pereboom's cases Plum's actions are causally determined by factors in over which he had no control, and it may even *seem* like this is a plausible explanation, the fact that this feature is shared among all cases is not enough to prove that it is in fact the *best* explanation for what drives intuitions. Instead, we must consider alternative explanations for what motivates intuitions. Therefore, it is important to investigate what influences our intuitions and correctly ascertain what *in fact* motivates our intuitions.

In addition to the methodological reasons I cite above for supposing that the four-case manipulation argument hinges crucially on a premise about what best explains our intuitions, this

is evidently also how Pereboom suggests that his argument should be understood. In a footnote, Pereboom states,

Al Mele (2006) argues that a manipulation argument against compatibilism need not be cast as an argument to the best explanation. I doubt that this is so. True, the argument can be represented without a best-explanation premise, but such a representation will not reveal its real structure. By analogy, the teleological argument for God [sic] existence can be represented as a deductive argument, but its real structure is an argument to the best explanation for biological order in the universe. The fact that the real structure of a manipulation argument against compatibilism is an argument to the best explanation becomes clear when one considers compatibilist objections to it—that, for, example, the non-responsibility intuitions can be accounted for by manipulation of a certain sort and not by causal determination. (2015, p. 79-80)

Here Pereboom makes it clear that his argument is one in which the explanation of intuitions is paramount. Furthermore, he states that the way one should object to his argument is by providing an alternative explanation for what else might explain our intuitions of Plum's non-responsibility. This is precisely what I have attempted to do in this thesis.

### ***5.3 Providing Evidence for the Best Explanation***

Insofar as we need to determine which features of Pereboom's four-case manipulation argument drive intuitions of non-responsibility, it is important to note that the claims Pereboom and myself make about what best explains intuitions are empirically testable claims. It's possible for instance to design empirical studies that manipulate the features of these cases in order to determine what does and does not motivate intuitions. Furthermore, we can test whether, after reading Pereboom's four-case argument, readers correctly understand the true causal nature of determinism. If it turned out that intuitions of Plum's non-responsibility are directly driven by the fact that Plum's actions are ultimately determined by factors outside his control, rather than any seemingly irrelevant features, and if after reading all four cases readers understand exactly what determinism entails,



then Pereboom's argument would successfully avoid my criticisms. I doubt, however, that this is what we would find.

One could simply manipulate the features within these cases and see if or how these changes affect intuitions regarding the extent to which Plum is morally responsible.<sup>13</sup> By individually removing features and aspects of these four cases that may drive intuitions, and testing whether Plum is intuitively morally responsible, we could tease out which features influence judgments and, therefore, what best explains why we have the intuition that Plum is not morally responsible.

Pereboom does not appear to be opposed to this type of investigation and remains confident that some of the factors I have discussed would turn out to be irrelevant to moral responsibility. Pereboom states, "If in these cases the manipulators were replaced by force fields or machines that randomly form in space that have the same deterministic effect on Plum as the manipulators do, the intuition that Plum is not morally responsible persists." (2014, p. 82) Therefore it seems Pereboom has no problem removing other features of the four-case argument to see if removing these features and changing aspects of the argument drastically alters intuitions of responsibility in the same fashion. Given this, I see no reason why Pereboom should object to a systematic investigation of this sort – one that involved changing different aspects of how these cases are presented – to determine whether Pereboom's explanation is indeed the best explanation for why we have the intuition that Plum lacks moral responsibility.

Before describing how, exactly, we might experimentally determine whether Pereboom offers the "best explanation" for intuitions, it is worth addressing the difficulty in determining what

---

<sup>13</sup> Here I am once again assuming first that the manipulation Pereboom describes is compatible with all the compatibilist requirements for moral responsibility, that readers find Plum intuitively not morally responsible, and that readers are properly understanding determinism and all of the relevant components of Pereboom's four cases.

“best explains” what drives intuitions of non-responsibility, given the fact that our intuitions are subject to so many factors. For instance, if one were to demonstrate that the order in which the cases are presented had *some* effect, it would be too quick to jump to the conclusion that order effects best explain what motivates our intuitions. Similarly, if we found that having one’s actions determined by factors outside their control significantly influenced intuitions, this would not be enough to conclude that this feature best explains why we have these intuitions either. Rather, it seems that the proper way to assess whether Pereboom is right about what *best* motivates intuitions of non-responsibility in Pereboom’s four cases would be to examine all of the features of these cases that do drive intuitions of non-responsibility and determine whether the intuitions are most significantly influenced by understanding that Plum’s actions are ultimately determined by factors outside of his control or if intuitions are more significantly altered by some combination of other factors that are independent of determinism, such as the order in which the cases are presented, the presence of agents intentionally manipulating Plum, or an emotional response to such manipulation.

In order to investigate to what extent intuitions are affected by different aspects of the four-case argument, we could slightly alter these cases and the way they are presented and measure how much these changes affect judgments of Plum’s non-responsibility. After presenting the cases independently to get a baseline for what intuitions readers have about these cases, we could present the cases unaltered in their regular order and in the reverse order (Case 4 first and Case 1 last) to see what effect order had on intuitions. Assuming that readers of at least some of these cases do in fact get the intuition that Plum lacks moral responsibility (and Pereboom’s manipulation cases satisfy Condition 2), we could next flip the order in which the information within each case is

presented (describe Plum's compatibilist capacities first and the manipulation last) and see if this has any effect on intuitions.

With this information available, we could then begin to slightly modify the cases to see what the intuitions actually track. For example, we could do as Mele suggests and present scenarios where determinism is false and agents have, to a minimal degree, the kind of free will that Pereboom thinks would allow for moral responsibility. If everything in the cases remained the same but Plum had some minimal sense of agent-causal free will (which Pereboom argues is necessary for moral responsibility) and yet readers have the intuition that Plum is not morally responsible, then we can conclude that it isn't the fact that Plum's actions are ultimately being determined by factors outside his control that motivates intuitions of non-responsibility. We could also replace the manipulators with force fields or machines, see if intent influences judgments, and even vary the extent to which it matters if the manipulators' intent matches the action that Plum performs. Of course, we would also need to verify that the subjects are not mistakenly attributing intent to these new manipulators and that their responses are not due merely to an earlier emotional response influencing judgments (Waller 2013; Weigmann et al. 2012).<sup>14</sup>

---

<sup>14</sup> Björnsson (in preparation) reports to have constructed a scenario where all compatibilist conditions for responsibility are met but where the cause of manipulation is an infection instead of the result of agents and that this non-intentional manipulator undermined attributions of free will and moral responsibility to the same degree as intentional manipulation cases. This study seems to be problematic for my account in that it suggests that agents and intent might not provide a better explanation for what motivates intuitions of non-responsibility in Pereboom's four-case argument. However, I am not necessarily arguing that agents or intent undermine responsibility. All I need for my argument to work is that the intuition that Plum is not morally responsible is the result of psychological influences such as order effects and not the result of readers recognizing that Plum's actions are causally determined by factors over which he ultimately has no control. I have previously discussed evidence which supports the claim that agential intent *does* influence intuitions of non-responsibility (Philips and Shaw forthcoming; Waller 2013). Providing evidence that intent *does not* best explain what motivates intuitions fails to demonstrate that features of determinism *do* motivate intuitions of non-responsibility. Nonetheless, in order for Pereboom's argument to work the latter must be the case.

Additionally, it would be useful to measure – and even *manipulate* – the readers’ emotional state before, during, and after reading each case, in order to ascertain whether having an emotional response of one kind or another influences judgments of responsibility, and whether emotionally responding to earlier cases influences later judgments. This has become a standard methodology in psychological studies on moral judgments. Furthermore, by presenting readers with different kinds of cases, one would know which features of these cases readers might be emotionally responding to (if any at all). Last, we could investigate whether readers understand determinism after reading the four-case argument – as Pereboom claims is the goal – or whether it in fact causes readers to become confused and assume that Plum’s compatibilist capacities are being bypassed or that they simply don’t exist. Either by asking follow-up questions or by asking for the reader to provide a written description of what occurred in the cases, we could investigate whether readers are conflating determinism with bypassing, or if they have some other problematic understanding that signifies that Pereboom’s manipulation argument fails to satisfy Condition 3 (see Murray and Nahmias 2014).

Noticing differences in intuitions as a result of altering the presentation of these cases would not be definitive proof that I am right. Pereboom and I agree that things like the order in which the cases will have an effect on intuitions and our understandings of the argument. If, however, intuitions were more significantly altered by a collection of factors that are not the product of understanding that Plum’s actions are ultimately determined by factors that are beyond his control, or if Pereboom’s presentation leads readers to erroneously understand Plum’s compatibilist capacities or the true nature of determinism, as I argue is likely the case, then the intuition that Plum lacks moral responsibility is better explained by these factors and Pereboom fails to provide the best explanation for what drives intuitions. Given the importance of correctly

explaining what drives intuitions for Pereboom's argument, by failing to consider these alternative explanations for what might motivate intuitions about his four cases, Pereboom fails to show that his explanation for intuitions is the best. If I am right and intuitions are being driven by alternative features of these cases, then Pereboom's argument fails.

## 6 CONCLUSION

The goal of this thesis was to demonstrate that arguments which depend on intuitions about thought experiments and hypothetical cases must offer a correct explanation for what features of the cases are motivating intuitions regarding the cases. I argued that psychological influences we are largely unaware of, such as order effects, can sometimes be significant enough to undermine arguments which employ these thought experiments. Without ensuring that our intuitions are tracking relevant features of an argument, intuitions regarding thought experiments will likely be unreliable and, therefore, fruitless for the purposes of certain kinds of philosophical debates. To exemplify these concerns, I presented Derk Pereboom's four-case manipulation argument. I provided evidence that suggests intuitions of moral responsibility in these four cases are better explained by order effects, likely driven by various irrelevant features such as agency detection, than by Pereboom's claim that the manipulation cases help readers recognize that Plum's actions are causally determined by factors outside of his control. It is possible Pereboom's argument may meet the first three conditions I argue are necessary to demonstrate a problem for compatibilism and thereby still succeed in providing a counterexample for the compatibilist account of moral responsibility. Nonetheless, if Pereboom failing to correctly identify what explains intuitions in the four-case manipulation argument, Pereboom positive argument for incompatibilism is unsuccessful. That is, if what leads one to judge Plum as not morally responsible is something that is not a feature of determinism (such as order effects that are motivated by an agency-detection

mechanism, an intent-tracking mechanism, or an emotional response to earlier cases), then it is not determinism that precludes moral responsibility in these cases and we have no reason to believe that determinism is incompatible with moral responsibility.

In response to those who might argue that order effects are intended to occur in Pereboom's four-case argument, I have argued that if the order effects are intended, then not only do features of determinism not provide the best explanation for what drives intuitions of non-responsibility, but that readers are likely being confused about what determinism entails. Furthermore, to those who object that proponents of manipulation cases do not require an explanation for what drives intuitions, in addition to demonstrating that Pereboom also understands his argument to require an explanation for intuitions, I have argued that the methodology employed by proponents of manipulation cases entails that what influences intuitions is relevant for moral responsibility. Therefore, insofar as a manipulation case is attempting to show that determinism precludes moral responsibility, we need to be sure that a feature of determinism (and not something else) is what motivates intuitions. If the intuitions are the result of other features or psychological influences, as I have suggested, then we lack evidence that determinism rules out moral responsibility.

My argument does not necessarily defend the compatibilist account of moral responsibility, nor does it suggest that our intuitions are too variable to be useful in philosophical study. Rather, I am providing reason for philosophers who rely on intuitions about thought experiments to consider the salience of subtle psychological influences when attempting to correctly explain what drives intuitions in thought experiments. One must take seriously the fact that our intuitions are influenced by many seemingly irrelevant factors when attempting to use thought experiments or hypothetical cases to provide support for an argument. Just as a good scientist considers all confounding variables before claiming to know the cause of a certain event, so too philosophers

must control for potential confounding factors in their hypothetical examples to ensure that they do not mis-identify what factors lead them to have the intuitions that they end up having.

## REFERENCES

- Adleberg, Toni, Morgan Thompson, and Eddy Nahmias. Forthcoming. "Do Men and Women Have Different Philosophical Intuitions? Further Data." *Philosophical Psychology*.
- Alexander, Joshua, Ronald Mallon, and Jonathan Weinberg. 2014. "Accentuate the Negative." In *Experimental Philosophy* (Vol. 2, p. 31). New York: Oxford University Press.
- Atran, Scott. 2006. "Religion's Innate Origins and Evolutionary Background." In P. Carruthers, S. Laurence, & S. Stich, S. (Eds.) *The Innate Mind: Culture and Cognition* 302-317 Oxford: Oxford University Press.
- Björnsson, Gunnar, Derk Pereboom. Forthcoming. "Traditional and Experimental Approaches to Free Will and Moral Responsibility." In Justin Sytsma & Wesley Buckwalter (Eds.) *Companion to Experimental Philosophy*, Blackwell
- Buckwalter, Wesley, and Stich, Stephen. 2011. "Gender and the Philosophy Club." *The Philosophers' Magazine*. (52): 60-65.
- Dennett, Daniel C. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: The MIT Press. 1984.
- Demetriou, Kristin. 2010. "The Soft-Line Solution to Pereboom's Four-Case Argument." *Australasian Journal of Philosophy*: 595-617.
- Feltz, Adam. 2013. "Pereboom and premises: Asking the right questions in the experimental philosophy of free will." *Consciousness and Cognition*, (22): 53-63.
- Fischer, John Martin, and Mark Ravizza 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68: 5–20.
- Giner-Sorolla, Roger; Russell, Pascale Sophie. 2011. "Moral anger, but not moral disgust, responds to intentionality." *Emotion*. (11): 233-240.
- Greene, J. D. 2011. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science*: 2105-2108.
- Guiseppe, Ugazio, C, Lamm, and T. Singer. 2012. "The role of emotions for moral judgments depends on the type of emotion and moral scenario." *Emotion*. 12(3): 579-590.
- Haidt, Jonathan. 2001. "The Emotional Dog And Its Rational Tail: A Social Intuitionist Approach To Moral Judgment." *Psychological Review*: 814-834.



- Haidt, Jonathan. 2003. "The moral emotions". In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. Oxford: Oxford University Press. 852-870.
- Haji, Ishtiyaque. 1998. *Moral Accountability*. New York: Oxford University Press.
- Haji, Ishtiyaque. 2009. *Incompatibilism's Allure: Principal Arguments for Incompatibilism*. Peterborough ON: Broadview Press.
- He, J.; X. Jin, M. Zhang, X. Huang, R. Shui, and M. Shen. 2013. "Anger and selective attention to reward and punish children." *Journal of Experimental Child Psychology* (115): 389-404.
- Hume, David. 1739/1978. *A Treatise of Human Nature*. Oxford: Oxford University Press.
- King, Laura A. and Hicks, Joshua A. 2011. "Subliminal mere exposure and explicit and implicit positive affective responses." *Cognition & Emotion*: 726-729.
- King, Matt. 2013. "The Problem With Manipulation." *Ethics* (24): 65-83.
- Levy, Neil. Unpublished manuscript. "Manipulating the Reader: Manipulation Arguments and Agency Detection"
- Li, Wen, Richard E. Zinbarg, Stephan G. Boehm, and Ken A. Paller. 2008. "Neural and Behavioral Evidence for Affective Priming from Unconsciously Perceived Emotional Facial Expressions and the Influence of Trait Anxiety." *Journal of Cognitive Neuroscience*: 95-107.
- McKenna, Michael. 2008. "A hard-line reply to Pereboom's four-case manipulation argument." *Philosophy and Phenomenological Research*, 77(1): 142-159.
- Mele, Alfred. 1995. *Autonomous Agents*. New York: Oxford University Press.
- Mele, Alfred. 2005. "A critique of Pereboom's 'four-case argument' for incompatibilism." *Analysis*: 75-80.
- Mele, Alfred. 2006. *Free Will and Luck*. New York: Oxford University Press.
- Mele, Alfred. 2008. "Manipulation, Compatibilism, and Moral Responsibility." *The Journal of Ethics*: 263-286.
- Mlodinow, Leonard. *Subliminal: how your unconscious mind rules your behavior*. New York: Pantheon Books, 2012.
- Mickelson, Kristin. 2015. "The Zygote Argument is invalid: Now what?" *Philosophical Studies*: 1-19.

- Murray, Dylan and Eddy Nahmias. 2014. "Explaining Away Incompatibilist Intuitions." *Philosophy and Phenomenological Research*: 434-467.
- Nichols, Shaun, and Joshua Knobe. 2007. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *NOUS*. 41(4): 663-685.
- Pereboom, Derk. 2005. "Defending Hard Incompatibilism." *Midwest Studies in Philosophy* 29: 228-247.
- Pereboom, Derk. *Living without free will*. Cambridge: Cambridge University Press. 2001.
- Pereboom, Derk. *Free will, agency, and meaning in life*. : Oxford University Press, 2014.
- Petrinovich, Lewis, and Patricia O'Neill. 1996. "Influence of Wording and Framing Effects on Moral Intuitions." *Ethology and Sociobiology* 17: 145-171.
- Phillips, Jonathan, and Shaw, Alex. .forthcoming. Manipulating Morality: Third-Party Intentions Alter Moral Judgments by Changing Causal Reasoning.
- Plaisier, Xanthe S., and Konijn, Elly A. 2013. "Rejected by peers—Attracted to antisocial media content: Rejection-based anger impairs moral judgment among adolescents." *Developmental Psychology*, 49(6): 1165-1173.
- Schwitzgebel, Eric, and Fiery Cushman. 2012. "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers." *Mind & Language*: 135-153.
- Schwitzgebel, Eric, and Fiery Cushman. Forthcoming. "Professional Philosophers' Susceptibility to Order Effects and Framing Effects in Evaluating Moral Dilemmas."
- Sinnott-Armstrong, Walter. 2008. "Framing Moral Intuition." In *Moral Psychology, Vol 2. The Cognitive Science of Morality: Intuition and Diversity*. 47–76. Cambridge, MA: MIT Press.
- Sinnott-Armstrong, W., Young, L., & Cushman, F. 2010. "Moral Intuitions" In *The Oxford Handbook of Moral Psychology* ed John Doris and the Moral Psychology Research Group. New York: Oxford University Press. 246-272.
- Sripada, Chandra. 2011. "What makes a manipulated agent unfree?" *Philosophy and Phenomenological Research*: 1-31.
- Tobia, K., Chapman, G., & Stich, S. 2013. "Cleanliness is Next to Morality, Even for Philosophers." *Journal of Consciousness Studies*, 20(11-12), 195-204.
- Todd, Patrick. 2012. "Defending (a modified version of) the Zygote Argument.

- Valdesolo, Piercarlo, and David DeSteno. 2006. "Manipulations Of Emotional Context Shape Moral Judgment." *Psychological Science* 17(6): 476-477.
- Vaesen, Krist, Martin Peterson, and Bart Van Bezooijen. 2013. "The Reliability of Armchair Intuitions." *Metaphilosophy*. 44(5): 559-578.
- Vihvelin, Kadri. "How Not to Think about Free Will." Kadri Vihvelin's Philosophy Blog. October 10, 2014. Accessed November 10, 2014. <http://vihvelin.com/>.
- Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Waller, Robyn. 2013. "The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument." *Philosophia*, (42): 209-222.
- Weigmann, Alex, Yasmina Okan, and Jonas Nagel. 2012. "Order effects in moral judgment." *Philosophical Psychology* 25(6): 813-836.
- Weinberg, Jonathan, Shaun Nichols, and Stephen Stich. 2001. "Normativity and epistemic intuitions." *Philosophical Topics*. 29(1-2): 429-460.
- Weinberg, Jonathan, Stacey Swain, and Joshua Alexander. 2008. "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp." *Philosophy and Phenomenological Research* 77(1): 138-55.