On Pereboom's Argument Against Moral Responsibility

Intro to Philosophy Paper #2

Ozaner Hansha

December 17, 2018

Abstract

This paper has 4 main goals: 1) Introduce the trichotomy of views that fall out of the irreconcilability of the free will thesis, determinism, and incompatibilism. 2) Extract Derk Pereboom's argument against moral responsibility from his article "Why We Have No Free Will and Can Live Without It"[1]. 3) Elaborate on Pereboom's justification of said argument. 4) Analyze and critique this argument and its associated justification.

1 The Trichotomy of Free Will

1.1 The 3 Propositions

Before we delve into Pereboom's argument, we need to clear up a couple of definitions:

- **The Free Will Thesis**: Some of the actions humans make are free. That is, while some of our actions may be forced or coerced in some way, there are certainly other actions that are done of our own accord.

- **Determinism**: Every event is caused by some prior event. That is, all events belong to an unbroken chain of events that would have invariably taken place given the same circumstances.
- **Incompatibilism**: Determinism implies that no action we make is free (i.e. determinism implies the free will thesis is false) and vice versa. In other words, incompatibilism asserts that free will and determinism are contradictory.

1.2 Justification

It would seem that, taken separately at least, all three of these propositions are reasonable:

- We generally believe ourselves as possessing the ability to make decisions of our own accord, that is we have **free will** (whatever we may define that as).
- Our intuitive understanding of cause and effect (as well as our training in classical physics) seem to imply that **determinism** is indeed the case.
- And finally, it would initially seem obvious that if all events are predetermined, including our own thoughts and actions, then we truly could not have a choice to make. That is, free will and determinism are **incompatible** and we must give up one or the other.

Notice that the last proposition, incompatibilism, makes clear that these three statements cannot be true simultaneously. The only way to deal with this situation, then, is to reject at least one of these statements. Below we will briefly define the 3 different views that correspond to rejecting each statement.

1.3 The 3 Stances on Free Will

- **Hard Determinism**: Accept determinism and incompatibilism, and thus reject the free will thesis. This means that no one ever does anything freely.
- **Libertatianism**: Accept the free will thesis and incompatibilism, and thus reject determinism. This means that, at least for free actions, some actions are not caused by previous events.
- **Soft Determinism**: Accept determinism and the free will thesis, and thus reject incompatabilism. This means that we can still act freely, despite all our actions having previous causes.

These are all the views that accept two of the statements and reject the other. However, there are other stances that may reject two or even all three of the statements. The **Hard Incompatabilist** view that Pereboom will push is one such stance that simply argues against compatibilist views (views that reject incompatabilism).

2 Extraction

2.1 Structure of the Arguement

The argument we are interested takes place in section 2 of Pereboom's article, titled "Against Compatibilism". Pereboom's argument comes in 4 cases, all of which he designs to satisfy the common conditions of compatibilist free will (second order desire, able to reason rationally/morally, etc.). The hope is that the cases will get progressively more realistic with the last case being a legitimate challenge to the notion of moral responsibility, which in turn challenges the compatibilist definitions of free will. At the same time, the argument depends on the idea that there is no real difference between any two consecutive cases that could allow one to create a new criterion for their preferred conditional definition of free will.

2.2 The 4 Cases

Case 1) Professor Plum is on the fence about killing Mrs. White. Even further he wants to want (has a second order desire) to do so. Moreover he can reason about the world rationally and understands morality. A team of scientists then send a radio wave to manipulate his brain and make his egoistic reasoning slightly stronger than his reason-responsive thinking (yet not so much so that it is inconsistent with his standard self) thus pushing him to kill Mrs. White.

Case 2) The same as Case 1 but instead of the scientists manipulating his mind right before the event, they instead manipulate him at the beginning of his life, which casually determines him to make that decision to kill Mrs. White.

Case 3) The same as Case 2 except now, instead of scientists doing this to him, it was the way his household and community raised and trained him as a child. He was much to young to prevent/resist this training.

Case 4) Plum is just a normal human like anyone else that satisfies the normal conditional definitions of free action, and decides to kill Mrs. White due to egoistic reasons.

2.3 The Argument

The argument can be formulated in the following way:

P1 Professor Plum is not morally responsible in case 1.

P2 There is no difference in moral responsibility between cases 1 and 2

P3 There is no difference in moral responsibility between cases 2 and 3

P4 There is no difference in moral responsibility between cases 3 and 4

P5 However, if Plum is not morally responsible in case 4, nobody can be morally responsible.

$\therefore$ Therefore, nobody can be morally responsible.

Note that the construction of the 4 cases depend on determinism being true, something Pereboom assumes in order to put it at odds with the conclusion. This will ultimately lead the reader, assuming they agree with the other premises, to have to choose between accepting determinism or the existence of moral responsibility.

3 Justification

3.1 Premise 1

Pereboom justifies this via the following intuition:

> *"...intuitively, he is not morally responsible for the murder, because his action is causally determined by what the neuroscientists do, which is beyond his control."*

And this, he assumes, is reasonable enough to agree with any standard notion of moral responsibility. Further, he considers the rebuttal that Plum may not be acting freely due to this being a temporally local manipulation of his brain state, but to that he counters:

> *"It is my sense that such a time lag, all by itself, would make no difference to whether an agent is responsible."*

Which again, is an intuitive enough proposition to accept. Why would free action, and thus the moral responsibility that comes with it, be dependent on how long ago another agent coerced you act in some way?

3.2 Premises 2,3 & 4

3.2.1 Case 1 to 2

Pereboom's justification of premise 2 is the same as his counter to the time delay rebuttal of premise 1:

"Again, it would seem unprincipled to claim that here, by contrast with Case 1, Plum is morally responsible because the length of time between the programming and the action..."

Thus reiterating that basing moral responsibility on an arbitrary time limit undermines it.

3.2.2 Case 2 to 3

Pereboom states that any challenge of this premise would have to show that Case 2 has some feature that morally distinguishes it from Case 3. He says there is no such feature since in both cases Plums action was determined by factors far in the past that were out of control. Whether or not it was done by his community and household or a team of neuroscientists is irrelevant, he argues.

3.2.3 Case 3 to 4

Again, Pereboom does not see any significant distinguishing feature between cases 3 and 4. He notes one difference is that, unlike in case 3, in case 4 Plum's actions were not caused by other agents (humans in this case) and instead by the chain of causality that envelopes all of our actions. He notes that this isn't morally distinguishing however because we can consider a case:

"...that is exactly the same as, say, Case 1 or Case 2, except that Plum's states are induced by a spontaneously generated machine-a machine that has no intellegent designer."

Here, Pereboom argues, Plum would still not be morally responsible.

3.3 Premise 5

It is at this premise that Pereboom reaps the reward of setting up the first 3 cases. Note that premises 1-4 imply that in case 4, Plum is not

morally responsible for the killing just as in case 1. But case 4, the most general case, encompasses all of human action. This is because if physicalist determinism is true, then all of Plum's (and anybody else's) actions would have been decided beforehand regardless of how they were raised or if anyone manipulated them directly. And this is just as true for those that do not satisfy the common compatibilist definitions of free action.

4 Analysis

Pereboom's argument strategy is quite effective as it assumes two totally reasonable assumptions (determinism and the fact Plum is not responsible in case 1) and shows it leads to the breakdown of moral responsibility and thus of the definitions of free action that supposedly defined it. He does this by introducing cases that, while staying morally equivalent to case 1, gradually look more like the general case of all human action.

Moreover, Pereboom not only showed that the most common compatibilist definitions of free will harbor this problem, but that any definition cannot suffice. He can guarantee this by making sure that there is no significant moral difference between each case. As long as this condition is held then there is no leeway for an opposing compatibilist to rebut with a new definition of what constitutes free action and thus moral responsibility.

Even further, the actual implementation of this gradual generalization is also quite effective. I am hard pressed to find a morally distinguishing feature that a compatibilist would be able to roll into a conditional definition of free action. Pereboom addresses the obvious differences between each case and shows that they are not significant to our purposes (when agents cause it, which agents cause it, and agents vs. non-agents).

5 Conclusion

5.1 Morality and Determinism

In sum, it would seem undeniable to me that determinism and a coherent definition of moral responsibility are irreconcilable. Indeed, we could make a case by case definition of morality that simply states Plum's hands are clean in case 1 but not in 4 but, this arbitrariness would go against our *other* intuitive notion of there existing a coherent *non*-arbitrary definition of morality. If there wasn't, why would we care for and abide by it?

5.2 Morality in General

Even further, it would appear to me that Pereboom's argument highlights the glaring problems in our notions of morality. Recall that Pereboom showed us case 1, a man being mind controlled to kill someone, is morally equivalent to case 4, a man killing someone in general. Any reasonable definition of morality, reasonable meaning agreeing with our intuition, would place case 1 squarely in the "not morally responsible" category whilst simultaneously putting case 4 in the "definitely morally responsible" category.

How then, can any consistent definition of morality exist that also conforms to our intuitive theory of morality whilst also conforming to our intuitive ideas of *how* such a theory should look like (i.e. that temporal/spatial proximity shouldn't matter, the identity of the agent shouldn't matter, etc.)?

5.3 Determinism

But to this idea, and to Pereboom's arguemnt, one may still object to the very construction of the cases themselves: "What if determinism is not true?". Rejecting determinism rather than accepting the non-existence of

moral responsibility is a valid interpretation of this argument, assuming its valid.

That said, while we cannot prove determinism, it seems far more intuitive and demonstrable than the very much known vagueness of morality. In fact, it is precisely this property that gives rise to moral dilemmas. If morality was founded upon clear and known principles then there would be no moral dilemmas. They would either reduce to mandatory decisions or coin flips.

Of course, while our most accurate physical models demonstrate just the opposite: that the universe is indeterministic, the indeterminism introduced by said theories (i.e. quantum wavefunction collapse) is not nearly sufficient to allow for theories of free will (i.e. free will collapses the particles in a certain manner or something to that effect) as that would violate their demonstrable randomness. Without free action, then, a theory of morality has to be based on something else and if not our intuition, what?

References

[1] Pereboom, Derk.

Living Without Free Will, Cambridge University Press, 2001