

Review: Derk Pereboom's Free Will, Agency and Meaning in Life.

Derk Pereboom's *Free Will, Agency and Meaning in Life* (Oxford, Oxford University Press, 2014, pp. 219) follows quite closely in the steps of his *Living Without Free Will* (2001). Pereboom argues for "hard incompatibilism" in both – the thesis that regardless of whether determinism is true, the kind of free will required for desert-entailing moral responsibility is unlikely to exist. In *Free Will...*, however, he pays more attention to issues other than moral responsibility. This, I think, is a great strength of the book. It has furthermore been suggested that *Free Will...* differs from *Living...* in being less radical; that Pereboom's hard incompatibilism has softened up over the years. I believe this view is mistaken. It is true that *Free Will...* contains several passages where Pereboom writes that he is happy to take on board important elements of moral responsibility theories usually labelled "compatibilist". However, *Living...* never states the contrary position. *Living*, rather, has a different focus – on other sceptical views, and how they differ from Pereboom's own.

Pereboom argues that we ought to give up angry blame and retributivism, but are justified in keeping many other practices sometimes believed to require free will. The latter makes Pereboom's hard incompatibilism easier to swallow, but it also creates a challenge. If X and Y are related practices, it is easier to argue that we ought to either keep or jettison both of them, than to argue that we ought to keep X while giving up Y.

Early in the book, Pereboom argues against libertarian views and compatibilism. Only the power of agent causation, according to Pereboom, could give us the kind of free will that can ground desert-entailing moral responsibility. He interestingly argues that agent causation is probably a coherent concept, but given what we know of science it is unlikely to exist. Pereboom also wants to defend a source incompatibilist view against a leeway version, and does so through the use of Frankfurt examples. Here, he runs into problems plaguing the entire Frankfurt debate. Early Frankfurt examples were criticized for either begging the question against incompatibilists or giving us a main character who could still do otherwise in important respects. In order to avoid these problems, Frankfurt examples evolved to become increasingly complex. Pereboom's own version, "tax evasion", is no exception. It is simple enough on the surface – Joe is a regular guy who decides to cheat on taxes for familiar reasons – but the reader is then given plenty of details about precisely how Joe's psychology works. I, for one, find it hard to generate any intuitions about cases as complex as this one. Ultimately, though, this is not important for Pereboom's larger project. Hard incompatibilism could very well be true even if moral responsibility requires leeway in addition to sourcehood.

When arguing against compatibilism, Pereboom revisits his famous four-stage manipulation argument. It features Professor Plum, who murders Ms White for selfish reasons, while supposedly fulfilling all compatibilist moral responsibility requirements. In the first stage, he is directly manipulated by neuro-scientists who can control his brain, and therefore not morally responsible. Pereboom's intention is to change only morally irrelevant features for each new step, forcing us to retain our initial judgement of non-

responsibility, until we arrive at stage four, ordinary determinism. In earlier versions of the argument, stage one featured a Plum who was remote-controlled from moment to moment by the scientists, but critics argued that such manipulation would rob Plum of real agency. In response, Pereboom changed stage one, so that the scientists merely press a button that momentarily enhances Plum's selfishness, similarly to how, e.g., hearing that your home team just lost might make you more selfish. The problem with this new version of stage one is that it is less obviously responsibility undermining. However, if I simply start with stage two, where Plum was programmed at birth to be the way he is, I still find the argument evoking strong incompatibilist intuitions, so this is a minor quibble.

The discussion about free will and rational deliberation is, overall, great. Incompatibilist philosophers have argued that it is irrational to deliberate about whether to do A or B, if you simultaneously hold the belief that which one you will do is already determined. Pereboom argues convincingly that deliberation can still be rational if the agent fulfils the twin conditions of epistemic openness and a belief in deliberative efficacy. The only thing I miss from this section of the book is a discussion of agents who regard it as settled that they will not A for *bad reasons*, and whose overall rationality would improve if they first changed their belief and then deliberated about whether to A after all. However, this is once again a minor quibble, since the chapter is already quite comprehensive.

The discussion about different ways of understanding moral responsibility and blame, and what is and what is not compatible with hard incompatibilism, is overall a really good one. Still, I believe there are some problems with the discussion of "ought" and free will. If "ought" implies "can", it is arguably the case

that no one ought to do otherwise than they do under determinism. Pereboom wants to resist this conclusion. In *Living...* he suggested that it is extremely counter intuitive that no one ever ought to do otherwise than they do, so it might be wiser to jettison the “‘ought’ implies ‘can’” principle than to accept this conclusion. However, this is a problematic line of argument for someone defending the perhaps equally counter intuitive conclusion that no one deserves to be blamed or praised for anything they do. In *Free Will...* Pereboom takes a different route, and distinguishes between the “ought of specific action demand” and the “ought of axiological recommendation”. Only the former is incompatible with determinism, according to Pereboom; the latter merely requires epistemic, but not metaphysical, openness. The problem with this strategy is that it remains unclear what the difference between these two “ought:s” really amounts to. They seem to play exactly the same role in deliberation, advice and planning, and so it remains unclear to me why the latter is compatible with determinism when the former is not.

Pereboom next discusses which moral responsibility practices we ought to give up. According to Pereboom, we ought to strive to eliminate expressions of resentment, indignation and moral anger when confronted with wrongdoing – although we can keep expressing disappointment, hurt, shock, moral concern and moral sadness. The reason we ought to stop, e.g., angrily yelling at wrongdoers, is that they do not *deserve* this and that doing so would be *unfair*. However, I believe that an analogous argument can be made for not expressing, e.g., sadness towards wrongdoers either. Being subjected to another person’s *sadness* over what one has done can be at least as painful as being subjected to her anger. If the agent could not help doing wrong, and if expressions of anger

would therefore be undeserved and unfair, so would (at least arguably) expressions of sadness. Thus, I believe that more arguments are needed for drawing the line where Pereboom does between justified and unjustified reactions to wrongdoing. Still, I believe he succeeds in painting a comprehensive picture of a possible alternative to current practices. And his subsequent call for reform of the criminal justice system is very convincing.

According to Pereboom, we ought to regard dangerous criminals as analogous to carriers of dangerous diseases. We have a right to protect society against the spread of dangerous diseases by quarantining people when necessary, and we likewise have a right to protect ourselves against crime by locking up criminals. Furthermore, still in line with the disease analogy, we ought to strive to *prevent* crime as much as possible, and rehabilitate the people we do lock up. Pereboom's view on criminal justice is far from a simple consequentialist one; he believes that the Kantian principle of not treating people as mere means to our ends is, even if not *the* supreme moral principle, a very important one. The criminal justice system must therefore be transparent and predictable, and the rights of the individual carefully weighed against the goal of protecting society. Pereboom furthermore writes that restorative justice processes could have an important place in the larger criminal justice system, and that milder punishments, like fines, can be justified on a deterrence basis (although deterrence alone cannot justify *incarceration*, since that would violate the Kantian principle). This latter claim is based on no more than Pereboom's intuitions, and so there is room for a bit more work to be done here. Still, the picture he paints of an ideal criminal justice system is a very attractive one – and I say this as a compatibilist. Pereboom actually presents a number of arguments

against retributivist criminal justice systems that are quite independent of the free will issue: Such systems are made to satisfy vengeful desires that ought not to be nurtured, they are terribly expensive, and it is doubtful whether it is really the state's job to make sure that people get what they deserve (if this really *were* a job for the state, we should also set up institutions designed to reward everyone who does a good deed). Add that a purely consequentialist deterrence system violates moral principles like the previously mentioned Kantian one, an objection which is also independent of the free will debate, and a compatibilist or libertarian might easily come to embrace Pereboom's proposed criminal justice system, although not for all of the same reasons as he does.

Towards the end of the book, Pereboom discusses the issue of whether we can have deep, loving, personal relationships if we do not believe that we have free will. It has been argued that this kind of relationship requires that you allow yourself to be vulnerable, and this in turn entails that you will feel resentment if the other party hurts you. Pereboom argues, on the contrary, that the necessary vulnerability could just as well be expressed through his preferred reactions of, e.g., sadness. To my mind, Pereboom makes a convincing case here. Even if it is psychologically hard (perhaps to the extent that we will never succeed *completely*) to give up resentment, indignation and moral anger, I believe Pereboom is right in claiming that having a close, loving relationship does not *entail* a susceptibility to these reactions when hurt.

Even though I find a line of reasoning here and there in the book to be weak or problematic, and even though reading the book did not really turn me into a hard incompatibilist, *Free Will...* moves me. It is not just the case that many arguments

and discussions in the book are thorough, thoughtful, clear and intuitively plausible. The entire picture that Pereboom paints in the second half of the book of a hard incompatibilist society is a highly attractive one. As I have already noted, I think Pereboom needs more arguments for drawing the line where he does between justified and unjustified reactions to wrongdoing. Still, Pereboom's vision of a world where everyone embraces hard incompatibilism can be summed up as everyone being much *nicer* than we are today. In this world, people are more understanding, more relaxed, friendlier. Whether he is ultimately right or wrong, there is something very attractive about this picture, and I cannot help being moved by it.

Overall, *Free will...* is an incredibly rich book, which I highly recommend to anyone interested in the debate about free will and moral responsibility, and who wants to get acquainted with some of the best arguments in the debate for a sceptical position.

Sofia Jeppsson, University of Gothenburg