Strawsonian Hard Determinism

Scott Hill

(This is a draft. Please cite the final version which is forthcoming in the Journal of Philosophy)

Abstract

Strawsonian accounts of moral responsibility are widely associated with opposition to hard determinism. However, it is only an historical accident that these views are bundled together. I show that Strawson’s deepest commitments are perfectly consistent with, and even support, a new and improved form of hard determinism. The resulting view is not revisionist about our practices in the way that extant versions of hard determinism are. After setting out my view, I then turn to Latham and Tierney’s (2022) objection to manipulation arguments. I make two points about their objection. First, while extant versions of hard determinism are vulnerable to Latham and Tierney’s objection, my version of hard determinism is not. Second, I argue that their objection raises important questions about what it is for two cases to be morally equivalent. I defend a principle about when cases are morally equivalent. My principle suggests a way to improve Latham and Tierney’s objection to manipulation arguments that are paired with extant variants of hard determinism. However, when paired with my version of hard determinism, manipulation arguments may accommodate Latham and Tierney’s objection.

Introduction: Strawsonian Hard determinism

The reactive attitudes include praise, blame, resentment, gratitude, forgiveness, anger, and love. Imagine we abandoned these attitudes. Much would be lost. Our interpersonal relationships would be destroyed. The richness and depth of human life would vanish, replaced with a mundane and empty existence. Furthermore, we are hardwired to praise and to blame and to love. Our lives are so constituted by such attitudes that it is difficult to imagine with any precision what we would look like, how we would be transformed, if the reactive attitudes were ripped from our mental and social lives. Perhaps, if those attitudes disappeared, we would disappear as well. Creatures without such attitudes, one might think, would not be us. These two ideas, that it would be bad to give up the reactive attitudes and that it would be difficult to give them up, are central to the Strawsonian account of moral responsibility.

Also central to Strawsonianism is the idea that moral responsibility is to be analyzed in terms of the reactive attitudes. What it is, for example, to be praiseworthy or blameworthy just is to be the appropriate target of praise or blame. Furthermore, Strawsonianism includes the idea that questions about whether a subject is morally responsible are settled entirely by our moral responsibility practices. If one wishes to settle questions about moral responsibility, one must not engage in esoteric metaphysical speculation divorced from our everyday practices. The place such questions are settled is in our practices. Look at the subject one wishes to evaluate. Measure that subject against our moral responsibility practices. And that is what will settle the issue of whether the subject is morally responsible. Thus, a typical form of Strawsonianism includes four claims:

S1: Giving up the reactive attitudes would be very bad because our interpersonal relationships depend on them.

S2: We are unable to give up the reactive attitudes.
S₁: Moral responsibility is constituted by being the appropriate target of the reactive attitudes.

S₂: Only considerations internal to our moral responsibility practices can threaten a subject’s moral responsibility.

Determinism is the view that the facts about the past, together with the laws of nature, entail every fact about the future. Take the state of the universe at the Big Bang. Add to it the laws of nature. Together they entail every fact about what happens after. This includes the fact that I type this sentence now. Hard Determinism is the view that Determinism is true and Determinism is incompatible with the existence of moral responsibility. So, for example, given that my typing now is entailed by the facts about the past and the laws of nature, I am not responsible for my typing now. And the same is true for any other action ever performed. From the most mundane actions to the most consequential and profound, no one is responsible for anything.

Strawsonianism is typically associated with the rejection of Hard Determinism. But I think it may instead be used to articulate and defend a new version of Hard Determinism. This view consists of the following claims:

SHD₁: S₁, S₂, S₃, and S₄ are true.

SHD₂: According to our internal practices of praising and blaming, determinism (or something close enough) eliminates moral responsibility.

SHD₃: Determinism (or something close enough) is true.

At the heart of the view I want to defend is the following idea: It may be bad to believe that we lack moral responsibility. That might destroy our interpersonal relationships. But distinguish believing that we lack moral responsibility from actually lacking moral responsibility. We can lack moral responsibility without believing that we lack it. The view I am defending is just that we lack moral responsibility. The Strawsonian considerations support the claim that believing we lack moral responsibility would destroy our interpersonal relationships. But they do not imply that merely lacking moral responsibility, without believing it, would require giving up the reactive attitudes. In making this distinction and employing it in this way, I am most centrally influenced by a tradition of thought running through Sorensen (1998) and culminating in Streumer (2013) and (2017). Though I discuss other influences below.

The Consistency of Strawsonianism and Hard Determinism
The idea that Strawsonianism conflicts with Hard Determinism is entrenched in the literature. If that is right, then my view would be mistaken. So I will say some things here to show that there is no such tension and that instead, every one of Strawson’s deep commitments fits very comfortably with Hard Determinism. It is only an historical accident that S₁ through S₄ come bundled with the rejection of Hard Determinism.

Hard Determinism, S₁, and S₂
There is a long tradition of error theorists about a domain defending their views from the claim that it has unacceptable consequences.

Normativity: The normative error theory is the view that normative judgments are beliefs that ascribe normative properties and that such properties do not exist. Some critics of the error theory, such as Dworkin
(1996) and Nagel (1997), argue that believing the error theory would lead to horrifying consequences. And that we should therefore reject the error theory.

Streumer (2013) and (2017) argues that we are incapable of believing the error theory. However, we can come close to believing it. We can believe different parts of it at different times. We can believe that various arguments for it seem sound. And although actually believing the error theory might have the horrifying consequences Dworkin and Nagel are worried about, coming close to believing the error theory in these ways, Streumer argues, would not. The undesirable consequences of believing the error theory and our inability to believe the error theory are compatible with the truth of the error theory as well as our coming close to believing that truth.

Truth: Alethic Nihilism is the view that nothing is true. Some critics argue that accepting this view would render us unable to express agreement. Gamester (forthcoming) argues that all that is needed to express agreement is to act as if there is a property that behaves in the way that truth is alleged to behave.

Mereology: Mereological Nihilism is the view that two objects never compose a third. Some critics of mereological nihilism maintain that accepting it would leave us unable to talk about artifacts. And that would lead to bad consequences. Van Inwagen (1990) argues that all that is needed to get the relevant benefits is to use artifact talk as paraphrase. We can talk about particles arranged pencil-wise rather than pencils, for example.

Some of the claims made by defenders of error theories about normativity, truth, and mereology parallel some of the claims Strawson makes about moral responsibility1. In each case it is maintained that it would be bad to accept an error theory. This parallels S1. In some cases it is maintained that we are not capable of believing the error theory. This parallels S2. However, the alleged practical consequences of accepting the relevant views can be avoided. In no case does the truth of the claims made by error theorists that parallel S1 and S2 require the falsity of the relevant error theory. In just the same way, I can happily accept Strawson’s S1 and S2 while consistently retaining Hard Determinism. It may be bad to believe that we lack moral responsibility. It may be impossible for us to do so. But that is all compatible with us actually lacking moral responsibility. And merely lacking it without believing that we lack it will avoid the destruction of our interpersonal relationships that Strawsonians fear.

Hard determinism, Todd, and S1
Patrick Todd (2016) has argued that S1 may be accepted by libertarians about free will. Consider again:

S1: Moral responsibility is constituted by being the appropriate target of the reactive attitudes.

Todd runs through the different ways of understanding S1 in the literature. Todd’s core idea is this: For each plausible way of articulating S1, the libertarian can just add that appropriateness is at least in part a function of the target’s actions being indeterministically caused. Nothing about S1, Todd shows, precludes such a way of understanding appropriateness.

I can incorporate Todd’s argument into the defense of my own view. My view is that no one is the appropriate target of the reactive attitudes. This is because determinism (or something close enough) is true and

---

1 Streumer (2017) discusses this sort of idea in connection with and defense of the younger Galen Strawson's regress argument against moral responsibility. To my knowledge, this idea has not been discussed in connection with the older P. F. Strawson’s views nor Pereboom's manipulation arguments.
determinism (or something close enough) rules out moral responsibility. So my view regarding \( S_3 \) is like Todd's with the addition that determinism (or something close enough) is true.

**Hard Determinism, Pereboom and \( S_4 \)**

Pereboom (2002) has argued that \( S_4 \) is compatible with the idea that determinism rules out moral responsibility. He points out that in general, theoretical considerations about metaphysics are relevant to practical matters about ethics. For example, our internal racist practices are sensitive to facts about whether the races are different. And learning about the metaphysics of race, that races aren't different, undermines the alleged justification to be racist. And so Pereboom thinks we shouldn't expect the reactive attitudes to be different in this regard. Just as general metaphysical issues are relevant to other practices, we should expect them to be internally relevant to our practices of praise and blame. Furthermore, Pereboom argues that cases of determination are sometimes relevant to our internal practices of praising and blaming. Internal to our practices, Pereboom maintains, is that finding out that a single individual was determined would undermine that individual's moral responsibility. So Pereboom thinks, finding out that everyone is determined should undermine everyone's moral responsibility.

Pereboom ends up with a kind of revisionism about our moral responsibility practices. There are things he thinks we can keep. But there is a lot that must be tossed out. In contrast, my view is that we don't have to toss anything out. We get to keep our practices. But I can nevertheless appropriate Perebooms insight that \( S_4 \) is compatible with Hard Determinism.

**But Responsibility Just Is Constituted By Our Responsibility Practices!**

I am arguing that Strawsonianism about moral responsibility is consistent with Hard Determinism about moral responsibility. Among other things, Strawson accepts:

\[ S_5: \text{Moral responsibility is constituted by being the appropriate target of the reactive attitudes.} \]

It seems to me that one could accept this claim. And one could accept that people are sometimes targets of the reactive attitudes. But one could also insist that no one is ever the *appropriate* target of the reactive attitudes. In that case, \( S_5 \) would be true. But no one would be morally responsible for anything.

The most frequent objection I have encountered is this:

For Strawson, moral responsibility *just is constituted by* our responsibility practices. So as long as I'm admitting the responsibility practices are there, that is all Strawson needs for there to be moral responsibility.

I think this is mistaken. Consider two examples: First, imagine someone accepted Adams’ (1973) modified divine command theory according to which:

\[ A: \text{An act is wrong iff that act is contrary to the commands of a loving god.} \]

Imagine, in addition, that such a person believed in Law’s (2010) evil god. The evil god is as similar to God as is consistent with being evil rather than perfectly good. So, among other things, the evil god is not loving.

If someone accepted Adams’ theory and also believed in the evil god, they would be an error theorist about wrongness. For Adams, not just any command constitutes an obligation. The command has to be issued by a *loving* god. It would not be plausible to maintain that an error theory is inconsistent with Adams'
theory because on Adams’ theory wrongness just is constituted by the commands of a god and there is a god. For Adams, the commands of a god are not enough. What is also needed is that the god issuing the commands be loving. If the god has issued the commands but isn’t loving, then there is no wrongness for Adams.

Second, consider Lewis’ (1980) account of pain according to which:

L: X is in pain iff X is in the state that occupies the causal role of pain for the appropriate population.

Lewis’ theory is consistent with eliminativism about pain. For example, imagine someone accepted L but didn’t believe in populations. Or alternatively, imagine they believed that there are populations but no populations are appropriate. For Lewis, it isn’t enough that someone is in a state that occupies a causal role. The state has to play the causal role of pain for the appropriate population. It would not be plausible to maintain that L is inconsistent with eliminativism about pain because for Lewis being in pain just is constituted by being in a state that occupies a causal role. If there are no populations or at least no appropriate populations, then there is no pain for Lewis.

In the same way, it is perfectly consistent to accept S3 and also be a hard determinist. For Strawson it isn’t just having the practices or being the target of such practices, by itself, that constitutes moral responsibility. The target has to be appropriate. For Strawson, the practices by themselves are not enough. If there are the practices but not the appropriateness, then there is no moral responsibility.

**Dropping the Appropriateness Condition**

I have been arguing that we can have our responsibility practices and that we can target people with the reactive attitudes even if we never do so appropriately. But the Strawsonian might insist that there is another way to interpret Strawson. One may leave out the appropriateness requirement and just maintain that to be morally responsible just is to be the target of the reactive attitudes. If we blame someone, then they are blameworthy. No need for talk of appropriateness.

If one drops the appropriateness condition, however, then it seems to me that Strawsonianism is subject to Euthyphro type objections. Very generally, such objections target theories according to which normativity depends exclusively on the contingent decisions or wills of an individual or group. The individual or group could make decisions or have wills that, given such theories, would make evil things good and wrong things right. In the present case, take the moral responsibility practices of the Nazis. Those practices targeted Jews as being blameworthy for various things. And yet, the Nazis were mistaken. The Jews were not blameworthy merely in virtue of the fact that the Nazis blamed them for things. This is similar to a point Todd (2016, p. 5) makes when investigating the possibility of dropping the appropriateness condition.

And so there is a dilemma for the Strawsonian. On the one hand, they may retain the appropriateness condition. In that case, it may be that people are targeted with the reactive attitudes but no one is appropriately targeted with the reactive attitudes. And so Strawsonianism is compatible with Hard Determinism. On the other hand, the Strawsonian may toss out the appropriateness condition. In that case, Strawsonianism is false. It is vulnerable to Euthyphro type objections.

**But the Rejection of Hard Determinism is Built Into Strawsonianism!**

Another suggestion one might have is that along with S1 through S9, Strawson accepts:

S10: Determinism is compatible with moral responsibility.
S₃ is part of the package of claims typically associated with Strawsonianism. If you get rid of S₃ and only keep S₁ through S₄, you no longer have Strawsonianism but something else. The rejection of Hard Determinism is built into Strawsonianism.

I am happy to give away the name ‘Strawsonianism’. I don’t care about names. What I care about is that the standard advertisement for Strawsonianism goes like this: Strawson has some deep insights. These insights include S₁ through S₄. In light of those insights, S₃ is true. My view is different. I grant that Strawson had some deep insights. But I deny that those insights establish S₃. Rather, as I will argue below, I think those insights affect the theoretical landscape in a way that is very different than what proponents of the standard advertisement for Strawson would have us believe. I think that S₁ through S₄ help make Hard Determinism plausible. And I don’t think they help the case for compatibilism at all. In short, if we define ‘Strawsonianism’ in the way that the objector suggests, then it remains true that none of Strawson’s deepest commitments are connected to the rejection of Hard Determinism. If we wish, we may tack on in addition to his deepest commitments that Hard Determinism is false. And we can use ‘Strawsonianism’ as the name for that combination of views. But it remains true that there is a minor variant of ‘Strawsonianism’, without the rejection of Hard Determinism artificially tacked on, and that only includes the claims that are unique to him, that fits with and (as I will argue below) even lends support to Hard Determinism.

Manipulation Arguments
Consider two cases:

*Existential Manipulation*: Plum decides to kill White because he is causally determined to do so by neuroscientists.

*Universal Determination*: Plum decides to kill White because he is causally determined to do so by the facts about the past and the laws of nature. Plum is an ordinary human being in this respect.

Cases like these are used to support Manipulation Arguments such as:

1. Plum is not morally responsible for killing White in *Existential Manipulation*.
2. If (1), then Plum is not morally responsible for killing White in *Universal Determination*.
3. So, Plum is not morally responsible for killing White in *Universal Determination*.

Premise (1) is motivated by intuition. A typical motivation for premise (2) will include carefully scrutinizing the differences between the two cases and arguing that none of those differences are morally relevant.

Latham and Tierney
Latham and Tierney (2022) show that, given Pereboom’s variant of Hard Determinism, premise (2) is false. In particular, accepting that Plum lacks moral responsibility in *Existential Manipulation* would allow us to preserve the reactive attitudes. Only Plum, in that case, lacks moral responsibility. So our practices will remain intact. On the other hand, accepting that Plum lacks moral responsibility for killing White in *Universal Determination* would force us to abandon the reactive attitudes. That would be very bad. That is a morally relevant difference. So (2) is false.

Strawsonian Hard Determinism and Latham and Tierney’s Objection
I accept Latham and Tierney's objection. I think it is sound. I think it is a successful response to the wide range of combinations of Hard Determinism and Manipulation Arguments that they consider.

But I think that when supplemented with Strawsonian Hard Determinism, the Manipulation Argument can withstand their objection. It may be bad to believe that we lack moral responsibility. But it may nevertheless be true that we lack moral responsibility. And even if we can't quite get ourselves to believe that it is true, we can come close to believing it is true. We can come close to believing we lack moral responsibility in exactly the way Streumer (2013) and (2017) says we can come close to believing the moral error theory. For example, we can believe different parts of it at different times. For another example, we can believe certain arguments for it seem sound. Nor do we even have to go that far. We might be able to come close to believing we lack moral responsibility in the sort of ways Streumer identifies. But it might be that, even if we lack moral responsibility, only a few strange philosophers are willing to come close to believing it. Most non-philosophers might be content never to consider the issue. And most philosophers might continue to maintain that those of us who deny we have moral responsibility are mistaken. All of this is compatible with our lacking moral responsibility. In light of these considerations, distinguish:

A: Believing that we lack moral responsibility would cause us to give up the reactive attitudes.

B: Lacking moral responsibility would require us to give up the reactive attitudes.

The upshot of manipulation arguments is that in Universal Determination Plum, and hence all of us, lack moral responsibility. The upshot is not that we believe Plum, and hence all of us, lack moral responsibility. All of the Strawsonian considerations support A. But it is B that would be required to identify a relevant difference between Existential Manipulation and Universal Determination. If we follow the advice of Pereboom, then we will indeed move from B to A. But we need not take Manipulation Arguments to support Pereboom's variant of Hard Determinism. We may instead take such arguments to support Strawsonian Hard Determinism. And in that case, we may agree that A is true. But we may see that B is false. Given Strawson's views of moral responsibility, we are unable to accept that we lack moral responsibility. So even if we lack it we will still go on believing we have it. And we will preserve the reactive attitudes. We will continue to have the rich, interpersonal lives that Strawson thinks are dependent on the reactive attitudes. It is just that no one is the appropriate target of such reactive attitudes. And so, given Strawsonian Hard Determinism, (2) is true. There is no relevant moral difference between Existential Manipulation and Universal Determination.

The Moral Equivalence of Cases Is Independent of Real World Consequences
What is it for two cases to be morally equivalent? Can the fact that what we believe about one case has different real world consequences than what we believe about another case be a relevant difference that by itself makes two cases morally different? Consider:

Red Cup: I took a sip of coffee from a red cup. I believe I took a sip.

Blue Cup: I took a sip of coffee from a blue cup. I believe I took a sip.

Imagine that my arm will be sawed off if I have a particular belief about one of these cases but not the other. For example, if I were to believe that there is no relevant moral difference between Red Cup and Blue Cup, then my arm will be sawed off. It seems to me that whatever the real world consequences of believing something about two cases are, such consequences are irrelevant to whether the cases are morally different.
For another illustration, consider Rachels’ (1986) examples below:

*Murder:* Jones will gain a large inheritance if his six-year-old cousin dies. One evening, while the child is taking a bath, Jones sneaks into the bathroom, drowns the child, and makes it look like an accident.

*Accident:* As before, Jones sneaks into the bathroom planning to drown the child. But, as Jones enters, the child hits his head and falls face down into the water. Jones stands ready to kill the child if necessary. But the child dies on his own.

Rachels claims that the two cases are morally equivalent. There is no relevant moral difference between the acts in the cases. But imagine someone objected to Rachels’ claim in the following way: Imagine they maintained that the consequences are not the same. While the consequences for the subjects within the relevant cases are the same, the consequences *for us here in the real world* are different. If we accept that Jones’ act in *Accident* is just as bad as his act in *Murder*, then it will make defenders of the distinction between killing and letting die sad. So if we accept that Jones’ act in *Murder* is bad to degree $n$, then no one will be sad. But if we accept that Jones’ act in *Accident* is also bad to degree $n$, then defenders of the distinction between killing and letting die will be sad. We don’t want to make anyone sad! So, the objector claims, there is a morally relevant difference between the cases. What we believe about those cases has very different consequences for us. It seems to me that Rachels should be unmoved by this sort of complaint. The real world consequences of what we believe about *Accident* and *Murder* are irrelevant to whether the cases themselves are morally different.

In light of examples like these, I endorse the following principle:

P: Any real world consequence of believing something about a case is not relevant to whether there is a moral difference between that case and another case.

Latham and Tierney’s objection to (2) is that accepting that Plum in *Universal Determination* lacks moral responsibility would require accepting that we lack moral responsibility here in the actual world. That would be a bad consequence for us here in the real world. However, accepting that Plum lacks moral responsibility in *Existential Manipulation* does not have that bad consequence. And so, Latham and Tierney maintain, there is a relevant moral difference between *Universal Determination* and *Existential Manipulation*. But given P, that is not a relevant difference. The fact that our interpersonal relationships would be destroyed if we were to accept a particular judgment about *Universal Determination* is not morally relevant.

**Moral Equivalence, Nearby Strawsonian Worlds, and Impossible Non-Strawsonian Worlds**

Latham and Tierney may improve their argument by maintaining that the consequences of accepting the relevant judgments about Plum do not merely apply to us here in the real world. Such consequences also apply to the inhabitants of Plum’s world in *Universal Determination*.

What they may say is this: If the inhabitants of Plum’s world in *Existential Manipulation* accept that Plum lacks moral responsibility, then they themselves can maintain their interpersonal relationships. But if the inhabitants of Plum’s world in *Universal Determination* accept that Plum lacks moral responsibility, then their interpersonal relationships will be destroyed. So that is a morally relevant difference between the two cases.

I think this is a plausible way to reformulate their argument so that it succeeds against Pereboom’s view. Here again, however, is where Strawsonian Hard Determinism shines. If we employ Strawsonian ideas such as $S_1$ and $S_2$ we can contrast two variants of *Universal Determination*.
Universal Determination (With Belief): Just like Universal Determination with the addition that everyone accepts that Plum lacks moral responsibility in virtue of being determined and every implication of accepting that.

Universal Determination (Without Belief): Just like Universal Determination with the addition that no one believes that Plum lacks moral responsibility in virtue of being determined.

We can see that Universal Determination (Without Belief) is to be preferred in running Manipulation Arguments. The key is to take onboard Strawsonian ideas such as $S_1$ and $S_2$. We can recognize that people are unable to believe that they lack moral responsibility. We can recognize that, even if they could, it would be bad for people to do so. Finally, we can appropriate the move error theorists have made in other domains: In particular, even if people are unable to believe they lack moral responsibility, they may nevertheless lack moral responsibility. If Strawson is right, and $S_2$ is true, then Universal Determination (With Belief) is a case that occurs in an impossible world. No one can believe that they lack moral responsibility. But in that case they do have the belief. On the other hand, Universal Determination (Without Belief) is a case that is clearly possible. To run the manipulation argument, we need not bother ourselves with cases taking place in impossible worlds in which people do things that are impossible for them to do. Better to have a case in which no one does anything that is impossible. Universal Determination (Without Belief) is such a case. If we judge that Plum lacks moral responsibility in Existential Manipulation, then the inhabitants of Plum’s world need not have any belief about the matter at all. Our judgment about whether they lack moral responsibility comes apart from any beliefs that they may have. In the same way, if we judge that Plum lacks moral responsibility in Universal Determination or its variants, the inhabitants of Plum’s world need not have any belief about the matter at all let alone beliefs that, given Strawsonianism, are impossible for them to have. It is not essential to the case that the inhabitants of Universal Determination believe that they, or that Plum, lack moral responsibility. Even if we accept that they lack moral responsibility, they may, and if Strawson is right they must, continue to believe that they have moral responsibility. And their valuable interpersonal relationships will remain intact.

And so we have a pair of cases, Existential Manipulation and Universal Determination (Without Belief) that are not different with respect to their consequences for their inhabitants. It is this pair of cases that should be used in the Manipulation Argument. And this Strawsonian variant of the argument may survive Latham and Tierney’s objection even if Pereboom’s variant does not.

References
Gamester W. Forthcoming Nothing is True Journal of Philosophy 1-25
Lewis D. 1980 Mad Pain and Martian Pain in Ned Block (ed.) Readings in the Philosophy of Psychology Harvard University Press
Nagel T. 1997 The Last Word Oxford University Press
Pereboom D. 2001 Living Without Freewill Cambridge University Press
Streumer B. 2017 Unbelievable Errors: An Error Theory About all Normative Judgments Oxford University Press
Van Inwagen 1990 Material Beings Cornell University Press

Acknowledgements
For comments and discussion I thank REDACTED.

Funding
Work on this paper was funded as part of the Euregio Interregional Project Network IPN 175 “Resilient Beliefs: Religion and Beyond”.