

Moral Praiseworthiness and Blameworthiness and the Free Will Debate

Conor Lutgen

Abstract

In this paper I aim to show that we do not have free will (regardless of the truth of determinism), and that we therefore do not have moral responsibility for our actions. However, I also argue that we can still live coherent moral lives in such a world, and that such a view does not commit me to skepticism about morality. I do this by distinguishing between moral responsibility and moral praiseworthiness/blameworthiness, and arguing that the latter is sufficient for moral action and evaluation.

I. Free Will and Determinism

The free will debate is a central topic in philosophy. The issue at hand is quite simple. On the one hand, we tend to consider ourselves as having free will as we go through our lives. We can make decisions and choices and act on them based merely upon what we want. But at the same time, we also tend to believe that the world is deterministic: that is, that the world is made up of a chain of events that causally relate to one another in the form of a necessary connection. But these two ideas seem to be at odds. If all things are causally determined, then why would we be any different? Are our actions simply the result of the circumstances around us?

This poses a dilemma. The problem can be stated like this:¹

¹ See Van Inwagen, 2008. While I support this argument (the Consequence Argument), I only agree with it insofar as it implies that determinism precludes moral responsibility—however, I argue that moral praiseworthiness and blameworthiness, as empirical descriptions about our beliefs, do not fall prey to the same problem as moral responsibility in this argument.

- 1) Our decisions are based on our personality and circumstances.
- 2) But our personality and circumstances are themselves determined by things beyond our control: namely, the past and the laws of nature.
- 3) Therefore, our decisions are based on things beyond our control.

(3) seems like a highly undesirable outcome. Not only does it imply that we lack free will, but it also implies that we lack any moral responsibility for our actions. This in turn means that no one can be said to have done something morally right or wrong, as they had no real agency in their action.

Philosophers have approached this issue in different ways. There are two main schools of thought: compatibilism and incompatibilism. Compatibilists hold that free will and determinism are reconcilable. Incompatibilists hold that they are not.

Of the incompatibilist group, there are two distinct camps. Determinists simply accept the above argument; determinism is true, and therefore we do not have free will. Libertarians argue that the world is not deterministic; in other words, that at least some events are *not* causally linked. Based on this, they argue that we do have free will.

There is also a viewpoint referred to as *free will skepticism*, or (similarly) *hard incompatibilism*. Those who hold this view argue that we do not have free will regardless of the truth of determinism.

In this paper, I present a somewhat different view. I hold that we do not have free will in either case (determinism or indeterminism), but also that this does not in fact threaten our practice of evaluating actions morally. In essence, I accept the argument, but I aim to show that the conclusion is not as bad as it seems.

II. Moral Responsibility

The first issue arises in defining moral responsibility. I advocate for a strong definition of the term:

Df. Moral Responsibility: For an agent to have moral responsibility for an act, it must be the case that (1) the agent is the ultimate source of that act, and (2) the agent wills the act.

This is equivalent to saying: for an agent to be morally responsible for an act, it must be the case that (1) and (2).

The key distinction made here is the “ultimate source” condition. This is a strong condition. It requires that the agent is not merely partially responsible for an act happening; and that the agent actually created that act. For example, if Smith shoves Jones and Jones falls into Brown, Jones is not responsible for the act of falling into Brown. For one thing, she was merely partially involved in the act—it would not have happened without Smith being there—and for another, she did not herself cause the act.

There can also be some confusion over what is meant by condition (2). This simply means that the act must be intentional; if the agent did not mean to perform the action, then it would be wrong to hold her morally responsible for it.

This is a fairly controversial way to define moral responsibility, as it means that someone who is only partially in control of an act is not morally responsible for that act. Others advocate a weaker view, saying for example that the agent must merely have caused the act (rather than being herself the source of the act) or else simply that the act must be such that we praise or blame the agent based upon its moral consequences. These sorts of definitions are ultimately unintuitive, I feel, and more or less amount to an attempt to shoehorn moral responsibility into a framework where it shouldn't fit. The latter definition is a fairly common one, and makes a

mistake in conflating moral praiseworthiness and blameworthiness with moral responsibility—something which I will focus on later in this paper.

The definition I have provided is, I believe, the best way to go about clarifying and defining the debate. This definition, or at least similar one, is certainly not unheard of in the free will debate; it is a position frequently held by those with incompatibilist views² for fairly obvious reasons. Using the same example as above, I can simply replace Smith with something like “the laws of nature and the events of the past” and we end up with an example in which Jones *does* in fact commit the act, but is not responsible because she was not the ultimate source for this—and this leads quickly to either skepticism about free will, or a libertarian view. Compatibilists will of course object to this definition; so I will attempt to provide some support in favor of it.

For one thing, the definition I stated has a strong intuitive appeal. Someone who is not well-read in the free will debate is likely to agree with it. It seems very natural to say that someone who is *not* fully responsible for her actions and who did *not* cause her actions is also *not* morally responsible for them. Abstracting away from the issue of determinism for a second, consider some basic examples.

- A) Jane is driving down the road. She suffers an unpredictable muscles spasm and swerves into an oncoming truck, resulting in some injuries. She may be held responsible in some sense—it’s certainly pragmatic to hold her financially responsible, for example, in order to uphold our legal conventions—but it doesn’t seem that she’s *morally* responsible. What happened was out of her control.
- B) Tom is mind-controlled by some sci-fi villain. Tom retains his identity to a reasonable extent; he’s unaware of the mind control until the villain triggers it and causes him to

² Cf. Kane, 2009

push an old lady down the stairs. Tom actually *did* the act, but he was not the ultimate source of it; rather, it was caused by the villain. Therefore, we don't hold him morally responsible.

C) Jed is an inept hacker. He attempts to steal a million dollars from charity, but instead accidentally donates a million dollars of his own money. While there might be some discussion as to whether this act was morally good or bad, it seems clear at the very least that he was not morally *responsible* for it, as he did not intend to perform the act.

D) Val is mugged while walking home. The mugger is fully in control of his mental faculties and forces her at gunpoint to give him her money. It seems that Val is *not* morally responsible for losing her money, and the mugger is morally responsible for robbing her, at least absent any considerations about determinism (however, I will argue that, once we include these considerations, the mugger is *not* morally responsible; at this point, I only want to motivate the intuitive view of moral responsibility).

All three of these examples describe how we generally tend to think about moral responsibility, and that sort of understanding would seem to support the definition that I gave. Ultimately, the central idea here is that we are not responsible for things that simply *happen to* us; we have to take an active role in order to be held morally responsible.

There is a further reason to prefer such a strong definition: it avoids arbitrariness. Any definition which is less demanding will have to answer a question as to *what degree* of control over an action is required for an agent to be held morally responsible for it. Consider a definition which modifies condition (1) to say merely that an agent must be *a* source for the act. Under

such a definition, there could be multiple sources causing an act; an agent could merely be part of a chain of events, yet still be responsible for the act. This certainly seems true in some cases.

Consider another example:

E) Bob curses loudly. This makes his friend Will angry—so angry, in fact, that he punches Bob and breaks his nose. Assume it is true that, had Bob not cursed, Will would not have punched Bob.

Here, it seems clear that Will is morally responsible, even though he is only part of a causal chain, and would thus not seem to be the ultimate source of his action. I would see this counterexample as an objection to my definition; but it can be easily answered.

First, an example like (E) is playing a bit of a trick by mixing up two contexts. On the one hand, it purports to describe moral responsibility on an intuitive level, away from the free will debate; but on the other hand, it also portrays the “ultimate source” condition on a level that is much more in-tune with that debate. The objection claims that Will is not the ultimate source of his actions, but is morally responsible, and therefore that condition (1) is wrong. However, if we consider Will’s actions on the same merely intuitive level on which we are considering moral responsibility, then it seems much more clearly to be the case that Will *is* the ultimate source of his actions. He seems able to do otherwise; he is not forced into this decision. Likewise, if we flip it around and consider moral responsibility in the more-complicated context of the free will debate, it seems likely that Will is *not* in fact morally responsible, as he was causally determined to act in the way he did.

I can also return to the original point about arbitrariness. The objector’s definition of moral responsibility will ultimately be arbitrary, which is a strong reason to buy my definition, even if we do consider it to have lost some of its intuitive appeal. In (A), for example, Jane is at

least *a* source for the action which occurs—the car crash—but she’s certainly not morally responsible. The objector must therefore qualify his condition, perhaps by saying that the agent must be the *primary* source of her actions. This will ultimately only lead to a problem of iterating vagueness. I can ask: at what point will Jane become the primary source of her action? Suppose she had medication that she could have taken to avoid the spasm; or suppose she knew she shouldn’t be driving due to the risk of spasms; and so on and so forth. The more suppositions I add on, the more the objector has to double-down and clarify his definition, to the point that one straw breaks the metaphorical camel’s back; one tiny adjustment makes the difference between moral responsibility and no moral responsibility.

My definition avoids this. “Ultimate” is an absolute, and thus does not suffer from any arbitrariness.

If we accept this definition, however, it seems that we’re committed to a harshly incompatibilist view. This means that we must either turn to libertarianism or free will skepticism, both of which are very polarizing, and, in different ways, rather unappealing. Instead, I’d like to offer another distinction, one which helps to clarify a great degree of the disagreement in the debate.

III. Moral Praiseworthiness and Blameworthiness (MP/B)

These terms are often brought up in the debate, but rarely given as much focus as moral responsibility; instead, they’re must often lumped in as a means of describing moral responsibility. However, distinguishing between moral praiseworthiness and blameworthiness (which I’ll be abbreviating as MP/B) and moral responsibility offers a great deal more clarity, and also a potential for reconciling our intuitive notions of free will and determinism with the

logical contradictions that they present together (without relying on implausible views of free will, moral responsibility, and so on).

The distinction is ultimately very simple. To say that someone has moral responsibility for an action is making a rather strong metaphysical claim; we are making a point about their causal interaction with the universe and the moral consequences of that interaction. To say that an action is morally praiseworthy or blameworthy is simply to say that we have a (presumably justified) attitude about its moral standing. It is thus an empirical claim. As such, we can define MP/B as:

Df. Moral Praiseworthiness: An act is morally praiseworthy if and only if we ought to praise the agent who did the act.

Df. Moral Blameworthiness: An act is morally blameworthy if and only if it we ought to blame the agent who did the act.

“The agent” might be a bit of a sticking point, particularly if we accept determinism. There is some desire to resist assigning an act to a specific agent if the doing of that act was never really up to her—but remember, this is an empirical claim. Thus, in this context, “the agent” is simply the person to whom we are assigning the praise or blame—not necessarily the entity that is ultimately responsible.

The real question is exactly what is entailed by the “ought.” The exact details will ultimately depend upon the moral framework we endorse. A moral constructivist would say that this means an act is praiseworthy if it upholds our society’s moral rules; a consequentialist would say that an act is praiseworthy if praising it has positive consequences; a deontologist would say that an act is praiseworthy if praising the act is a good thing in itself. For the purposes of this

paper, I'll remain agnostic as to which moral theory we should use; that's a topic for ethicists. An intuitive idea of morality should be sufficient for this paper.

This distinction may seem trivial, and it may seem almost indistinguishable from the idea of moral rightness and wrongness. The crucial difference, though, is that MP/B makes a claim about the attitudes that we should have about an act, and *not* about the act itself. Conceivably there is a case wherein an *act* might be morally right, but it would be wrong to praise it. This puts MP/B in a different position in the dialectic; in a sense, it is more of a pragmatic claim than other moral claims are. And, most importantly, its requirements are weaker. It need not be the case that an agent is ultimately responsible for an act in order for it to be the case that we ought to praise that act. In this sense, it is an empirical claim. An act is praiseworthy insofar as praising it is a good thing—for example, insofar as praising it gives us positive consequences (for a consequentialist).

With MP/B distinguished in this way, we have two different axes of moral evaluation at play in the free will debate (with responsibility being the other). This allows for some more refined arguments, but it also necessarily complicates the dialectic, especially considering how often MP/B and responsibility are tied together; as such, it falls on me to give some reasons for making this distinction, outside of merely proving my own points.

Firstly, the distinction allows us to discuss complicated examples with more clarity. Consider this case:

- F) George is an old man who lives alone. Concerned about his safety, he buys a gun to keep in the house in case there's ever an emergency. He is proficient in its operation, but otherwise not particularly knowledgeable about it. One night, he has some friends over. One friend, Cosmo, puts a bullet in the chamber without George knowing.

Later, George is showing another friend the gun and, after checking that the magazine is out, pulls the trigger. The gun fires and wounds his friend.

This is a difficult situation to evaluate. On the one hand, George probably ought to have been more careful. He probably ought to have learned more about gun safety, and known to check the cartridge for a bullet. He also probably should have avoided a situation where the gun was out and pointing at someone. But is he *responsible*? That's difficult to answer. There were certainly epistemic barriers here; yet, at the same time, it would have been very easy to avoid the situation wherein his friend was hurt. Moreover, Cosmo clearly had something to do with the injury—but does that erase George's responsibility? Likewise, is Cosmo now *himself* responsible for the injury, despite never pulling the trigger?

If we introduce MP/B as well, however, there's a lot more clarity in the discussion. It's fairly evident, for example, that Cosmo is at least morally blameworthy; yet, assuming that he didn't intend any harm to come from his acts, he doesn't seem responsible. He certainly wasn't the source of the act in question. Likewise, we can certainly say that George deserves some blame just for pulling the trigger—that was a blameworthy act. We may still remain uncertain as to the question of responsibility, but we've come a lot farther having introduced MP/B than we could with only moral responsibility. These sorts of situations are much more easily understood with the combination of MP/B and moral responsibility.

Secondly—and perhaps more importantly—MP/B allows us real license to sidestep the difficult issues of causation that the free will debate often ignores. There is a quagmire of skeptical problems that arises when we begin to consider any sort of causation in detail, and for the most part, the free will debate has done its best to avoid stepping foot in any of those issues. A Humean can shed doubt on almost any view of free will simply by asking how we know that

causation even exists. After all, or so the claim would go, isn't our only evidence for causation inductive, and thus fallible?

In the past, the response has generally been to bracket these sorts of concerns for others to discuss. We're generally convinced that causation exists, even if it's difficult for us to *prove* that it does. But that response is a bit unsatisfactory, and it seems to me that the dialectic on this matter falls into what is more or less a state of mutually assured destruction—I won't bring up Hume if you don't bring up Hume, because neither of us want to (or can) answer him. But at the same time, how can a debate that is so heavily analytical, and so intensely involved in the concept of causation, fail to answer these questions? Sure, Humean skepticism may be a bit beyond the reach of this debate; but there are some issues that aren't. How, for example, can I even define one *action*, when all actions take place in an endlessly complicated causal chain? Is my action pulling the trigger, or is it shooting the gun? It seems impossible to non-arbitrarily define a single event, and then assign that event to an actor. Sure, I may perceive the event as "George pulled the trigger;" but what real *reason* is there not to see this event as six instances of "George pulled back the trigger one-sixth of the distance needed to cause the gun to shoot?" And of course that problem iterates—one-twelfth the distance, one-one-hundredth, and so on. This poses a difficult question for anyone trying to develop a coherent account of determinism or indeterminism; these accounts rely *heavily* on the idea of certain events causing (or not causing) other events. How can I be deterministically caused to steal if the act of stealing is not separable from all the circumstances which are supposed to have caused it? How can I be an agent-cause of my own action if my action has no definite beginning?

Talking about actions in terms of MP/B allows us to more or less ignore this problem. There doesn't need to be some actual, metaphysically delineated event to which we are

responding; it merely needs to be the case that we can have an praise or blame something about what it is that we perceived. As such, so long as we all more or less agree on how we perceived the event beginning and ending—which we most often do—there is no reason that we can't assign praise or blame. Essentially, MP/B shifts the field of inquiry from the metaphysical to the empirical. It needn't be the case that some external state of affairs is true (for example, *George is morally responsible*) in order for us to make claims about MP/B. Because it operates in this realm, we are far more justified in avoiding those skeptical concerns.

Now, of course, it may be the case that this is the realm of discussion which the free will debate is already in; but if that *is* the case, it's best for us to be explicit about it, and to adjust our terminology accordingly. Thus I think it's fair to at least say that we ought to adopt the moral responsibility/praiseworthiness and blameworthiness distinction, regardless of the rest of my argument.

IV. Motivating Hard Incompatibilism

With all this in mind, I put forth the following claim: we are never morally responsible for our actions. Making this argument requires both that I give evidence for its veracity and that I motivate it in a way that meshes with our intuitions about morality and free will. In approaching that first hurdle, I will be presenting a view that aligns strongly with Derk Pereboom's hard incompatibilist viewpoint³. However, we differ significantly on both what the impacts of such a view are and on the strength of the view. For me, this is a much weaker framework than it is for Pereboom. His view seems to eschew not only moral responsibility, but also MP/B; I only doubt the former.

³ Pereboom, 2006; Pereboom, 2009.

The argument is very straightforward:

- (1) If determinism is true, then we do not have moral responsibility.
- (2) If indeterminism is true, then we do not have moral responsibility.
- (3) Either determinism is true or indeterminism is true.
- (4) Therefore, we do not have moral responsibility.

Essentially, I aim to shed doubt on both compatibilism and libertarianism as viable accounts of moral responsibility.

A. Determinism Precludes Moral Responsibility

Here, I aim to debunk the compatibilist claim that determinism and moral responsibility are compatible. Generally, the compatibilist makes this claim by rejecting the Principle of Alternative Possibilities (PAP), which states that an agent is morally responsible for an act only if she could have done otherwise.⁴

This principle represents our intuitive notion of free will as it relates to moral responsibility, and while it is contested, it still receives strong support upholding it. Generally, the compatibilist response can be grounded in Frankfurt-style cases (named for the philosopher who first give such an example). The case goes something like this:

- G) Black, a neuroscientist, puts a chip in Jones's brain that allows him to control Jones's actions. Jones is a member of a jury deciding on the outcome of a trial. Black wants a guilty verdict. If it seems that Jones is about to render a verdict of not guilty, Black will intervene and force Jones to render the guilty verdict instead. As it happens, Jones renders a guilty verdict on his own.

⁴ Cf. Fischer, 2010; Fischer et. al., 2010.

This is meant to show that we can be morally responsible for our actions despite the inability to do otherwise. In (G), Jones could not have done otherwise than render a guilty verdict—Black would have prevented it—but we still consider him to be responsible for the verdict he gave because Black did not need to intervene at all.

Incompatibilists object by posing a dilemma.⁵ In the example given, it must be the case that the world is either deterministic or indeterministic. If the world is deterministic, then Jones never could have done otherwise than to render a guilty verdict. The example tries to substitute *determinism* with Black, but Black would still exist within the context of a world where determinism is true, making Black's role redundant. On the other hand, if the world is indeterministic, then Jones's actions will not be the result of a causal chain, making it impossible for Black to effectively intervene; it will be impossible for him to know which verdict Jones will render before Jones actually renders the verdict.

Compatibilists can formulate replies for both horns. On the indeterministic side, they modify the example. Suppose that there is some prior sign that is a necessary, but not sufficient, condition for Jones rendering a verdict of not guilty. If he is to render the not guilty verdict, he will first roll his eyes, for example. He might also roll his eyes and render a guilty verdict—his actions are still not causally determined. But at the same time, Black can now effectively intervene. This reply seems plausible, but it's unclear whether or not it gets around the initial issue. Black's intervention still relies upon the sign having some kind of deterministic connection to the action—while Jones' eye-rolling might not require a deterministic link to the verdict, it is still the case that the verdict *cannot* occur without the prior sign, which implies some kind of deterministic link; and, arguably, Black needs determinism to be true in order to intervene at all,

⁵ Kane, 1985; Widerker, 1995.

as such intervention is itself a causal relation between Black's action and the chip in Jones' brain.

The deterministic horn is trickier for the compatibilist to deal with; however, compatibilists like John Fischer have claimed⁶ that the Frankfurt case is at least showing that *if* causal determinism does indeed rule out moral responsibility, it is *not* by way of denying alternative possibilities. Rather, the *combination* of a deterministic world and Black's intervention is what rules out moral responsibility in this case.⁷

On either horn of the dilemma, however, the compatibilist is ultimately conflating moral responsibility with MP/B. I would argue that, assuming the Frankfurt case works, it still does not show that Jones is morally responsible for his actions, but rather that he is merely blameworthy. The example seems to be relying on the same intuitive notion that Jones's being the source of his action makes him morally responsible for it. The argument doesn't seem to be advocating for a *different* view of moral responsibility—it's merely showing how the existing view might coexist without alternative possibilities. But that's not the crux of the issue. The problem is not that determinism itself rules out alternative possibilities, but that it rules out the very ability to *be* the source of one's own actions. Given the fixity of the past and the laws of nature, it is impossible for an agent to independently choose and enact her actions. The Frankfurt case doesn't do anything to deal with that issue.

Indeed, this sort of case is more of a motivator for the MP/B distinction than anything. The metaphysics of the world determine whether or not Jones can be the ultimate source of his own actions. Black merely determines whether or not we can hold him blameworthy of praiseworthy. If Jones makes the decision on his own, then it seems that he is praiseworthy or

⁶ Fischer, 2010

⁷ There are potential problems with this claim, as well; see Garnett, 2013.

blameworthy, despite lacking alternative possibilities. But this is not a particularly strong claim; consider this example:

H) Alicia is the best sprinter in the world—so good, in fact, that she cannot do other than win every race she runs. No one else can come close to beating her, and she's psychologically incapable of doing anything other than running as fast as she can in each race. She also knows that her winning inspires many other athletes around the world, so that they also improve and are happier. Similarly, she knows that almost no one roots against her; her winning doesn't bring people pain.

When Alicia wins a race, that act is praiseworthy because it brought happiness to a great number of people and improved their lives. Even if we know that she simply cannot do other than run in and win the race, we still consider winning a morally praiseworthy deed. It would be odd to think that her inability to do otherwise somehow precludes us from giving praise; indeed, that would rather be a further reason *to* praise her. We can think similarly about someone who is incapable of doing a bad deed, or someone who is incapable of doing a good deed. In both cases, the fact that the person's psyche removes any alternative possibilities does nothing to eliminate their praiseworthiness or blameworthiness.

This response raises another sort of question, however: if I'm willing to concede that MP/B exist without alternative possibilities, then am I really rejecting compatibilism? Is this argument at all substantive?

The short answer is *yes*. By rejecting compatibilism, I don't mean to make as strong a claim as most incompatibilists; indeed, I'd much rather my position remain attractive to compatibilists as well. However, there is a substantive difference between a deterministic world with moral responsibility and one with only MP/B. The compatibilist claim implies that we are

capable of being the ultimate source of our own actions. We can be held fully morally culpable for their consequences, and we in a sense active; we have agency and we can be held accountable for that (and based upon it).

A deterministic world with *only* MP/B does not imply this same source of agency. We are merely reactive; what we do is not up to us. Any moral focus on our actions is *necessarily* grounded in practicality, and not on a cosmic ideal of the moral good or moral bad. This sounds dour, but in the end such a world seems to align much more closely with the one we actually live in, and seems far more intuitively plausible.

B. Indeterminism Precludes Moral Responsibility

Here, I respond to the libertarian claim that we have moral responsibility only in an indeterministic world. There are two objections that have been raised to this claim: the luck objection⁸ and the disappearing agent objection.⁹

The luck objection is quite straightforward. If it is the case that our actions are not part of a causal chain, then it seems to be merely a matter of luck that I do one action and not another. This seems to undermine moral responsibility as it implies that our actions are not under our control.

The libertarian has a response to this. Kane¹⁰ offers an example that runs like this:

- I) An assassin is trying to kill the Prime Minister. Due to shifting wind patterns, there is only a 50% chance that his bullet will hit. By luck, it does.

⁸ Pereboom, 2009.

⁹ Pereboom, 2012.

¹⁰ Kane, 2009.

Here, the success of the assassin's action is merely a matter of luck; but do we consider him any less responsible for having killed the Prime Minister? Kane thinks the clear answer is *no*. Kane gives several examples like this one, and urges us to look at the indeterminacy as an obstacle which we have to overcome; for any action we take, there is some degree of indeterministic "noise" which we have to get through in order for the action to take effect.

However, framing these examples differently makes them much more shaky. Instead of (I), consider this example:

- J) The same assassin is trying to decide whether or not he wants to take a contract to kill the President. He has strong reasons to take the contract and strong reasons not to; but by luck, he decides not to.

Here, the assassin's responsibility is much more questionable. The issue is that he doesn't seem active in his decision, as Kane would like him to be. Kane wants the assassin to *decide to decide* not to take the contract; and then he faces indeterministic noise, and then his decision carries through. But in reality, it seems as though the indeterminacy comes in at a higher level—at the level of *which way* the assassin decides to decide. As such, I find Kane's reply to the luck objection largely unconvincing.

This example also relates nicely to the disappearing agent objection. This objection voices a similar worry to the one I just brought up: if it is the case that events are undetermined, we can imagine scenarios in which an agent is perfectly torn on which decision to make. In such a situation, the issue is not even that luck would decide which action ends up happening, but simply that *there would be no* deciding factor as to what happens. The agent is thus wholly uninvolved in the decision, which once again seems to undermine moral responsibility.

Kane formulates a view of libertarianism that attempts to reply to these sorts of objections by allowing for some amount of causal connection. Not *every* event needs to be causally indeterministic, Kane says; merely *some* must be. In such indeterministic events, we may perform *self-forming acts*, actions which are uncaused but which then causally determine our personality and how we act in future scenarios. Thus we have agency in *developing* the ways in which we respond to different events—even events in which we are causally determined to act in a certain way (or at least highly probable to act in a certain way).

This notion of self-forming acts does something to counter the objections with luck and disappearing agency; however, I find this account highly implausible. For one thing, the constitution of a self-forming act—what it is that *makes* an act self-forming—is unclear and arbitrary. It seems that the only criteria are that the act must be a torn decision, and that it must somehow impact the agent's character. This seems far from a robust account of free will; almost any act can reasonably be seen to have *some* impact on one's character. It seems likely that the only truly torn decisions are minute ones, like picking a pair of socks in the morning or deciding between pasta and pizza for lunch. In such cases, the stakes of our decisions are very low, which makes us more willing to decide without considering all factors. Thus, it would be more likely that we do simply and indeterministically decide; our actions would appear to be less clearly a result of mere circumstance. We are much more reasons-responsive when dealing with decisions of greater importance. We take more time and look carefully at each option; and such deliberation is therefore far more likely to be influenced by factors entirely outside our control, rather than purely internal factors.

Moreover, these self-forming acts have their own problems with disappearing agency. In such a torn decision, there are two options: either we are equally likely to decide either way (a

50/50 split) or we tend toward one option (a 60/40 split, for example). In the former case, it seems that we have very little control over which decision we make—and then our self-forming acts are themselves merely products of luck, which seems like a poor way to ground moral responsibility. But on the other hand, the more we tend toward one decision, the more problematic each option becomes. Suppose I have an 80% chance of choosing to do A and a 20% chance of choosing to do B. If I do A, then my action veers much closer to determinism—the more often I do the higher-probability act, the closer to determinism I am. Then it seems that I am less free in a sense; my actions are all-but causally determined. Of course, I still have the option to do B instead; but if I do B, then I seem to have lost control over my act. I was foiled by the indeterministic “noise.” Either way, it seems as if I am lacking the sort of control required for a robust account of moral responsibility.

There are also issues with the very notion of self-forming acts. They’re intuitively unappealing; they don’t seem to describe any real phenomenon, but rather create a phenomenon in order to support the libertarian theory. There’s a problem with regress, as well. When I’m deciding which decision to make in a self-forming act, my decision is influenced by my past self-forming acts—a chain which ends in a decision that was based either in a causally determined background or in pure luck, neither of which is appealing. Furthermore, there is a potential worry with second-order decisions; that is, while deciding about my self-forming act, must also decide to decide, and so on? It seems that I would have to do so to be seen as active in the way that Kane imagines it—battling against the indeterministic “noise.” But this would result in an infinite regress.

As such, this account of libertarianism seems highly unconvincing.

Some have also advocated a different kind of libertarianism: agent-causal libertarianism. In such a view, the agent as a substance has the power to simply act. Agents are unmoved movers; they can cause things without themselves being caused by things.

This view has several issues, perhaps the biggest one being its physical implausibility. Given our best physical theories, it seems unlikely that a being could exist such that some part of it is outside the realm of causality. The appeal to “agents” is also problematic. It’s highly arbitrary, and any definition of an agent in this sense is likely to be implausible. Being a member of the human species seems like an arbitrary criterion for agent-hood. Being rational is less arbitrary, but runs into issues with marginal cases—where exactly is the line to be drawn? Can someone have a low enough IQ to lose out on their agent status? Do infants miraculously achieve free will upon reaching some level of rationality? Most criteria for this status will likely run into similar problems. Finally, there is something of an intuitive lack of appeal to this claim. When we make decisions, it doesn’t seem as if we are somehow separate from the world of causality; indeed, in order for us to be reasons-responsive, we must be causally influenced to some degree. As such, this libertarian view also looks to be highly implausible.

C. We Lack Moral Responsibility

Given the problems that both determinism and indeterminism pose for moral responsibility, it looks reasonable to conclude that we lack moral responsibility in any case. This conclusion is not a particularly appealing one on its surface. It runs counter to our desire to view ourselves as free agents who are responsible for our own actions. It may be much less dour than it seems, however—an outlook which I will hope to motivate in the following sections.

However, this sort of skepticism about moral responsibility is ultimately more plausible than any of its rival views. We tend to resist this outlook, but it's difficult to deny the intuitive logical appeal. Moral responsibility is a strong requirement, and once we begin looking more deeply at causal connections, it seems impossibly demanding. In almost any case, if we add enough details, we can begin to agree that the agent is not morally responsible. A murderer would seem to be morally responsible on the face of it; but what happens when we learn about his various psychological defects, his childhood traumas, his abusive parents, and the harsh environment he grew up in? The more we understand the factors that led to any act, the less likely we are to hold someone morally responsible; and this isn't a bad thing. This sort of skepticism about moral responsibility can be useful in our practical considerations, and it makes for a more realistic picture of morality.

V. Morality Without Moral Responsibility

We tend to resist skepticism about moral responsibility on the basis that it more or less renders the field of ethics null and void. But this is not the case. The lack of moral responsibility is only problematic if it significantly impedes our ability for moral judgment and action; here, I aim to show that it does not do that. The basis for this claim is quite simple: MP/B is sufficient for both moral evaluation and moral action.

The general assumption is that moral responsibility is necessary for moral evaluation. I don't find this to be the case, for a reason that may seem backwards to some: *if* we assume that we don't have moral responsibility, it does *not* follow that we lose our capacity for moral evaluation.¹¹ If tomorrow we were given definitive proof that we have no moral responsibility

¹¹ Others have argued similarly. Cf. Strawson, 1962.

for any of our actions, we would not then start saying that none of our actions have any moral status. We would still see murder as being wrong and giving to charity as being good. We may change our views on some things, such as punishment for moral wrongdoing, but we would not assume that morality is now an empty construct.

This claim may seem somehow illicit. After all, it does not provide positive reasons for believing that we *do* have moral evaluation without moral responsibility; it merely postulates an alternative in which that *would* be the case. That is a legitimate concern, but ultimately unwarranted. For one thing, the assumption to which I am replying is very clear: moral responsibility is necessary for moral evaluation. If we can conceive of a case in which we can have moral evaluation but not moral responsibility—as, indeed, I have just shown—then the assumption is not sound.

Moreover, however, the reason I don't see this claim as being worrying is because it simply *is* the world we live in (at least according to the view which I am trying to defend). This makes it all but impossible to give an example without crossing contexts. If I put forward a case where we clearly don't have moral responsibility—such as the mind-control case in example (B)—then it is also the case that the agent in question shouldn't receive moral credit or blame. In that example, Tom's action (likely) isn't morally wrong; it's the villain's action (the mind control) that is wrong. If I give an example like (D) (the mugging), where an act *is* clearly morally wrong, then the immediate intuition is to say that the agent has moral responsibility.

To give a positive example, then, I have to modify the case to also include details about determinism that exclude the possibility of moral responsibility. Let's say I add these details to D and call the case D*: Val is mugged and the mugger is in full control of his faculties, *and* the world is either deterministic or indeterministic in such a way that the mugger is not the ultimate

source of his actions. Now we have a very clear case where the agent does *not* have moral responsibility, but where his action is nonetheless obviously morally wrong. This is because the act can be morally blameworthy despite the lack of moral responsibility.

This in turn is simply because the requirements of MP/B are much weaker than those of moral responsibility. It only needs to be the case that we are justified in assigning moral blame to the mugger. That is clearly true. The justification can come from a number of sources—the merely practical, for one. It would be ineffective to *not* assign him blame for his action. We wouldn't be upholding the moral standards of our society, and upholding these standards makes life a lot easier for everyone. But the justification also comes from moral sources. If we say that the act *just is* wrong, then it also seems clear that assigning moral blame to the act *just is* right—even though the agent is not morally responsible. It may be that he was simply incapable of doing anything other than committing the mugging; yet that doesn't mean that we can't blame him. Consider example (H) again about the sprinter who cannot fail to win. There, it seemed very clear that she still deserved praise for her victory. The same logic applies here.

It is thus possible to morally evaluate an action even without moral responsibility. Moreover, I'd like to reinforce the agnosticism of this approach to any particular moral theory. The approach has constructivist undertones, which may drive some away; it seems to imply that all there is to morality is our constructed notion of it. But this is so. It still can be (and likely is) the case that there is an objective moral right and moral wrong—these would simply tell us whether or not we ought to praise or blame an act. The only claim I am committed to is that MP/B is compatible with an absence of moral responsibility for the act.

This framework is also sufficient for us to *act* morally. If I do an action that is morally good, then I am morally praised for it—even though I was not the ultimate source of that action.

I am still acting within the sphere of ethics and morality simply by doing something which has moral consequences.

There are potential counterexamples to be raised. For instance: a natural disaster also has moral consequences, but it seems wrong to hold it morally blameworthy. Certainly a hurricane causes plenty of pain and long-reaching loss; but is it then acting within the sphere of morality, just as I am? But this case is relevantly distinct. The natural world is not an agent; thus, it cannot receive moral praise or blame. We can see something similar in a case like (B), the mind control example. The critic can provide such an example wherein the mind-controller is made analogous to determinism. In that counterexample, it seems obvious that we cannot blame Tom for his actions while being mind-controlled; this would cast doubt on the idea that we can blame people in a deterministic world. But determinism is relevantly distinct to the mind controller. The past and the laws of nature are not agents; they cannot be assigned moral blame. The mind controller *is* an agent; he can be assigned moral blame, and thus we assign it to him. But we do not then re-allocate that blame upon learning that the mind controller was merely subject to the whims of determinism; the latter is not an agent, and so the chain of blame ends. This response might beg a question—the critic can use an example like the following:

K) Elaine fires an employee who did not deserve to be fired. However, she only did this because her own boss threatened to fire *her* unless she fired someone else; and her boss only did that because her friend Jerry sneezed on him; and Jerry only sneezed because he had a cold that he caught from a waiter.

Now, is Elaine blameworthy for her act, or have I backed myself into a corner? Must we move the blame down the chain of agents until it terminates, and end up assigning the blame to the poor waiter? Surely this would be absurd.

The answer is straightforward. Blameworthiness requires that we have justification for assigning blame—that we *ought* to do it. It *may* be the case that *some* of the blame can be transferred along the chain, proportionally to the amount of justification at each level; but that's not particularly absurd. We would still leave Elaine with most of the blame, but it seems acceptable to put some of the blame on her boss, and some of the blame on Jerry, and perhaps a tiny amount of blame on the waiter. The critic could argue that I am then committed to saying something similar about the mind control case—that is, that Tom deserves some of the blame for his action. This critic's claim would be dubious (as the justification for assigning blame is very slim), but I'm ultimately willing to bite the bullet. In fact, it seems to align fairly well with our intuitions to give some tiny amount of blame to Tom, if a case could be made that we ought to do so.

I take myself at this point to have shown that MP/B is perfectly adequate for moral evaluation and action; but I'd like to go further and also point out that it is in many ways a preferable framework for prescriptive ethics. It aligns well with our moral intuitions, and allows us to articulate a good deal more about cases where we may find the agent innocent, but still wish to say that his action was wrong. Consider again example (C) about the inept hacker who donates to charity by accident. There, we can say that the act was morally good, but perhaps not morally praiseworthy, as we ought not give the hacker praise for an accidental act of good. As such, I think that this framework not only provides a perfectly suitable basis for morality, but is also successful in helping to address some of the more difficult moral issues that ethicists often struggle with. This constitutes a further reason to accept the MP/B framework, even absent the consideration of the free will debate.

VI. Consequences for Practical Life

The next question for me to consider is a fairly pressing one for any compromise position: *so what?* I have thus far laid out a framework which seems to merely say “everyone is wrong, but only slightly.” It seems that it would be very easy for philosophers with almost any other viewpoint to simply accept this framework into their argument without drastically changing their views—so why should it matter?

There are several reasons. The first and most immediately relevant is that it greatly clarifies the debate. If we are to have differing views, we should at least be clear on *which points* we differ on. Ultimately, the goal is indeed to remove opposition, rather than create it. Hopefully this view is palatable enough to at least partially depolarize the various viewpoints.

But more importantly, my viewpoint actually carries with it practical implications for our lives. In this way it is unlike both compatibilism (which more often aims to simply reconcile the views we already have, rather than describing a way in which our views are wrong) and incompatibilism (which requires either that determinism or indeterminism is true—a condition that is all but impossible to determine empirically). My view describes the notions we have, but also the confusion regarding those notions, and as such has some practical impact on the way we look at the world.

The largest impact has to do with criminal punishment. Skepticism about moral responsibility sheds some doubt on the way we deal with criminal justice presently. Pereboom, whose hard incompatibilist view is similar in this respect, offers a variety of ways in which our criminal justice system can be reconciled with a lack of moral responsibility.¹² It seems impossible—or at least, very difficult—to reconcile a retributivist justice system with a lack of

¹² Pereboom, 2006.

moral responsibility. An eye for an eye is not a good maxim for a world where no one has ultimate control over the eyes they're taking.

Pereboom's preferred strategy for dealing with this problem is arguing by analogy with a quarantine. In a quarantine, certain individuals are contained to mitigate a threat to society, even though those individuals are not themselves responsible for the threat they represent. Criminals could be treated the same way. This analogy is fairly effective, but it does entail a large paradigm shift in terms of our treatment of criminals. There ought to be a much greater focus on rehabilitation, and prisons ought to be made as comfortable as reasonably possible (in much the same way that quarantine zones are not deliberately uncomfortable). It would also require that we do as much as we can to prevent individuals from becoming criminals—this means greater investment into at-risk areas, and a greater focus on providing non-criminal opportunities to potential criminals.

Most of these changes sound like good ideas; however, there is a certain degree of intuitive opposition to Pereboom's approach. For one thing, we resent the idea of particularly heinous criminals living in comfort. This can be partly justified practically, as a deterrent; but that justification leads to a whole host of other problems. If punishment is merely meant as a deterrent, then it ought to be as painful as possible (even inhumane), and it's not necessary that those punished *be* guilty—only that they *appear* guilty to us. It seems more accurate to instead say that heinous criminals deserve to be blamed for their acts, and that this blame in turn provides some justification for punishment—or at least for *not* prioritizing comfort.

I would advocate for a slightly different approach to criminal justice. On one hand, my view would fully support Pereboom's ideas about preventative measures. If we accept that criminals are not morally responsible for their actions, then our focus must necessarily shift

toward preventing those actions in the most effective way, which would seem to entail removing the sorts of circumstances that generally lead toward criminal behavior.

However, the MP/B framework also allows for a moderately adjusted view of the criminal justice system. We can still assign moral praise and blame to individuals for their actions, but we can't hold them morally responsible. Blameworthiness seems to provide at least partial justification for incarceration and the like. We can say that, though a criminal may not be responsible, they still should not have done the act—or that we ought to blame them for that act. To this extent, we are at least justified in using punishment in terms of practical considerations, to reinforce that certain kinds of acts are undesirable. Combined with Pereboom's quarantine analogy, we have justification for a criminal justice system that is fairly intuitively valid.

Prevention is the most pressing factor, and, after that, punishments like incarceration can be justified to a reasonable extent. Allowing for blameworthiness at least reduces the obligation we have towards criminals' comfort, though we would likely still have much higher obligations in this regard than we consider ourselves to have currently. That said, the position would represent something of a compromise between Pereboom's position and the justice system we have today, and at least in my view, it aligns more closely with our intuitions about justice than either of those systems do.

Another practical concern that is often raised about the lack of moral responsibility is that it would dilute our ability to use moral language in conversation. If I'm not responsible for my actions, can someone say, for example, that I am a virtuous person, or that I have poor moral character?

I personally don't see such concerns as having much ground. For one thing, they seem to be approaching the problem in the wrong direction; even if there were a vast enough paradigm

shift to render these terms incoherent in the sense that we currently mean them, it is far more likely that our language would simply adapt naturally. The objection seems to rely on the implication of an absurd and undesirable consequence; but I don't think that's realistic.

There are other replies to this objection, though. Pereboom¹³ offers a very compelling response: we don't require responsibility in other, similar cases. For example, consider a man who is an excellent athlete. We praise him for being an excellent athlete. But really, he isn't the one who is responsible for becoming an excellent athlete. He was born with good genes that made him athletic and well-suited to the game, and his parents raised him to be determined and passionate and to keep training. But does that change our view that he's an excellent athlete? Not at all. It's still perfectly coherent to give him that sort of praise. The same can be said of moral language in conversation. To say "she's a virtuous person" doesn't imply "she is morally responsible for becoming a virtuous person;" it's simply an empirical statement.

That isn't to say that our personal relations would be completely unaffected if everyone were to accept this view, but that fact hardly constitutes an objection. For one, it simply won't be the case that everyone—or even a decent number of people—accept this (or likely any other) philosophical view of free will. And even if they did, the change would be mostly positive; the assumption that circumstances cause our behavior tends to breed empathy more than anything.

Thus, this framework does very little to threaten our conception of morality in everyday life, and actually entails a fairly appealing paradigm shift.

¹³ Fischer et. al., 2010.

VII. Conclusion

By introducing the MP/B distinction, we can better describe and understand the issue of moral responsibility and determinism. I have given several independent reasons to accept this distinction; but I've also shown how it can be used to create an attractive compromise position within the free will debate, one which also aligns closely with many of our intuitions about free will and moral responsibility.

Works Cited

Fischer, John Martin (2010). "The Frankfurt cases: The moral of the stories." *Philosophical Review* 119 (3):315-336.

Fischer, John Martin, et al. *Four Views on Free Will*. Blackwell Publishing, 2010.

Garnett, Michael. "Fischer-Style Compatibilism." *OUP Academic*, Oxford University Press, 23 Feb. 2013, academic.oup.com/analysis/articleabstract/73/2/387/170392?redirectedFrom=fulltext.

Kane, Robert (2009). Libertarianism. *Philosophical Studies* 144 (1):35-44.

Kane, R. 1985. *Free Will and Values*. Albany: State University of New York Press.

Pereboom, Derk. *Living without Free Will*. Cambridge Univ. Press, 2006.

Pereboom, Derk. "The Disappearing Agent Objection to Event-Causal Libertarianism." *Philosophical Studies*, 2012. <https://link-springer-com.colorado.idm.oclc.org/article/10.1007%2Fs11098-012-9899-2>.

Pereboom, Derk (2009). Hard incompatibilism and its rivals. *Philosophical Studies* 144 (1):21 –

33. <https://link-springer-com.colorado.idm.oclc.org/article/10.1007%2Fs11098-009-9371-0>

Strawson, Peter F. (1962). Freedom and resentment. In Gary Watson (ed.), *Proceedings of the British Academy*, Volume 48: 1962_. Oup Oxford. pp. 1-25.

van Inwagen, Peter (2008). The Consequence Argument. In Peter Van Inwagen & Dean W. Zimmerman (eds.), *Metaphysics: The Big Questions*. Blackwell.

Widerker, D. 1995. Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *Philosophical Review* 104: 247-61.