Mitigating Soft Compatibilism

Justin A. Capes

Florida State University

Final version published in *Philosophy and Phenomenological Research* 87: 640-663

Let *compatibilism* name the thesis that determinism is compatible with free action and moral responsibility.[1] Compatibilists are united by their belief that determinism does not preclude moral responsibility or free action. They are divided, however, on whether the truth of determinism is nevertheless somehow relevant to moral responsibility. Two main positions on the issue have emerged. There are those who say that the truth of determinism is relevant because it is required for moral responsibility, and those who deny this and who insist instead that the truth of determinism is entirely irrelevant to the issue of responsibility. Most compatibilists fall into one or the other of these two camps. But a third position is available that, with one exception to be discussed towards the end of the paper, has yet to be seriously explored by compatibilists and their opponents, according to which determinism *mitigates* (without necessarily eliminating) responsibility.[2] My principal aim in what follows is to explicate this novel brand of compatibilism and to highlight the unique resources it possesses for dealing with a popular and forceful style of argument against compatibilism. I also discuss one of the most pressing challenges facing a compatibilist view of this sort, and offer some

---

[1] Determinism is the thesis that, at any instant *t*, only one future is compatible with the state of the universe at *t* and the laws of nature. Some theorists (see e.g., Clarke 2003: 4) understand determinism to include the additional thesis that every event (with the possible exception of events occurring at the first instant, if there was a first instant) is deterministically caused by prior events. I will do so as well.

[2] Yet another position, which I will not explore here, is that determinism *enhances* responsibility.

suggestions as to how proponents of the view might attempt to address that challenge. Attention to the first two compatibilist positions adumbrated above will help set the stage for the subsequent discussion.

1. Hard and Soft Compatibilism

According to what I will refer to as *hard* compatibilism, determinism is compatible with, and, indeed, is required for, free action and moral responsibility. The most demanding version of this view says that an agent *S* performs an action *A* freely and is morally responsible for *A*-ing only if determinism is true. A slightly less demanding version allows for free action and moral responsibility in indeterministic universes, but nevertheless insists that there must be a deterministic connection between pertinent psychological features of the agent and the agent's behavior. This less demanding version recognizes that deterministic causal relations can obtain even in universes that are not entirely deterministic.[3]

Hard compatibilism is not as popular as it once was. Nowadays many theorists, including many compatibilists, think that even the less demanding version of the view demands a bit too much. Those who are sanguine about the prospects for free action and moral responsibility in suitably indeterministic contexts believe that hard compatibilism imposes an exceedingly stringent requirement on these things and is thus to be rejected. Hard compatibilism is, moreover, inconsistent with one popular idea motivating recent compatibilist accounts of moral responsibility and free will.

---

[3] Versions of hard compatibilism have been defended by Ayer 1954, Hobart 1934, Nowell-Smith 1948 and Smart 1961, among others.

Consider the following story. Ann, a normal, morally competent adult human being, intentionally kills her husband, Bob, despite knowing that it is morally wrong to do so. This action of Ann's was indeterministically and non-deviantly caused by pertinent beliefs, desires, and intentions of hers—for example, her belief that shooting Bob would result in his death, her desire to rid the world of him and her intention to do so straightaway—and there are no obvious exculpating considerations. Ann was in her right mind at the time, she wasn't manipulated or coerced, she did exactly what she wanted to do and, indeed, exactly what she wanted to want to do, she identifies with and wholeheartedly endorses the desires that motivated her murderous action, she has and regularly exercises powers of self-control, has the ability to regulate her behavior in accordance with both prudential and moral reasons, and could have done otherwise than kill Bob (on both a compatibilist and incompatibilist reading of "could have done otherwise"). Those who think agent-causation is possible and that it is required for acting freely and being morally responsible may suppose that Ann agent-caused her action.

Did Ann freely shoot and kill Bob, and is she morally responsible for killing him? Hard compatibilists, it seems, must answer *no*. Because Ann's behavior is not deterministically caused by pertinent psychological states and events, both varieties of hard compatibilism identified above imply that Ann did not kill Bob freely and, consequently, that she is not morally responsible for killing him. However, many theorists—compatibilists and incompatibilists alike—will no doubt find this result difficult to accept. Ann satisfies a robust set of conditions that a majority of contemporary compatibilists and many incompatibilists would deem sufficient for acting freely and being morally responsible. These theorists will find it hard to believe that,

despite satisfying those conditions, Ann gets off the hook simply because the causal connection between her action and its immediate psychological antecedents was not deterministic. A hard compatibilist might try to convince them otherwise, for example, by offering an argument to the effect that indeterminism brings with it too much chance or luck to accommodate free action and moral responsibility, but they are not persuaded by any extant argument of that sort, and neither am I.

That Ann did not act freely and is therefore not morally responsible for what she did owing significantly to the lack of appropriate deterministic causal connections will be especially difficult to believe for those who accept the idea that our most basic views about ourselves as persons, including our view of ourselves as free agents who are morally responsible for at least some of what we do, should not be held hostage to the details of our best scientific theories. I will henceforth refer to this idea as *the resiliency intuition*.[4] The resiliency intuition has motivated important recent work on the topics of free action and moral responsibility. For instance, John Martin Fischer, a compatibilist, is explicit about the role this idea plays in his own work on these topics. He tells us that he is motivated to defend compatibilism in large part because he believes that "our basic status as distinctively free and morally responsible agents should not depend on the arcane ruminations—and deliverances—of the theoretical physicists and cosmologists" (2006: 5; see also Fischer and Ravizza 1998: 16).

Incompatibilists insist that whether anyone ever acts freely or is morally responsible hinges on whether determinism is true. Evidently, then, they must reject the resiliency intuition. Notice, however, that hard compatibilists are no better situated with respect to the resiliency intuition than incompatibilists. On their view, whether anyone

---

[4] Here I follow Dan Speak 2008: 124.

ever acts freely or is morally responsible depends on whether the relevant deterministic connections obtain no less than it does on an incompatibilist view of free action and moral responsibility. Should it turn out that our actions are indeterministically caused, hard compatibilism would have the unsavory consequence that no one ever acts freely or is morally responsible. Like incompatibilism, then, hard compatibilism seems to make our status as free and responsible agents contingent upon the deliverances of our best physical theories. Compatibilists like Fischer who are motivated by the resiliency intuition are therefore likely to eschew hard compatibilism in favor of a less demanding position, according to which free action and moral responsibility are compatible with the truth of both determinism and indeterminism. I will refer to this view as *soft* compatibilism. According to soft compatibilists, the truth or falsity of determinism is entirely irrelevant to whether anyone ever acts freely or is morally responsible.

We can distinguish between two varieties of soft compatibilism. *Standard* soft compatibilists say not only that the truth of determinism is irrelevant to whether anyone ever acts freely or is morally responsible, but also that the truth or falsity of determinism has no bearing whatsoever on the issue of moral responsibility. However, this further claim is not entailed by soft compatibilism. Soft compatibilists needn't insist that determinism is entirely irrelevant to the issue of moral responsibility. It is open to them to say, instead, that while the truth of determinism does not preclude the sort of freedom or control minimally required for some degree of moral responsibility and is therefore irrelevant to whether we are free agents who are responsible for at least some of what we do, being deterministically caused to act by factors over which one has no control and for

which one is not responsible may nevertheless mitigate one's moral responsibility. This is *mitigating* soft compatibilism. Let's take a closer look at it.

2. The Contours of Mitigating Soft Compatibilism

To begin to understand mitigating soft compatibilism, it will be helpful to distinguish two questions we might ask about an agent's responsibility for a morally wrong action *A*.

Q1: Is *S* to blame, and does *S* deserve punishment, for *A*-ing?

Q2: How much blame, and how severe a punishment, does *S* deserve for *A*-ing?

Obviously, a negative answer to the first question implies a negative answer to the second; if the answer to *Q1* is "no," then the answer to *Q2* will clearly be "zero blame and zero punishment." However, an affirmative answer to *Q1* does not tell us much about how we should answer *Q2*, as there may be various aggravating or mitigating factors that need to be taken into consideration.

It is the position of mitigating soft compatibilists that the truth or falsity of determinism by itself (i.e., apart from other responsibility-subverting considerations) has no bearing on the answer to *Q1* but that it does have a bearing on the answer to *Q2*. According to them, being deterministically caused to perform a morally wrong action by factors beyond one's control and for which one is not responsible is a mitigating factor; it diminishes (without necessarily eliminating) one's blameworthiness and, accordingly, the severity of punishment one may properly be said to deserve. One of their central thoughts

is that, although determinism is compatible with the strongest sort of freedom or control minimally required for some degree of moral responsibility, agents in certain indeterministic contexts may exercise more robust forms of control over their behavior and, for that reason, may be deserving of more blame and harsher punishment for their bad actions than their counterparts in deterministic universes.[5]

The mitigating soft compatibilist is thus resolute in her compatibilism; she denies that determinism by itself precludes moral responsibility and, indeed, with one possible exception, insists that the truth or falsity of determinism is entirely irrelevant to the question of whether we are ever morally responsible. The possible exception is this.[6] Although all mitigating soft compatibilists will insist that determinism is consistent with responsibility, some proponents of the view might want to allow for the possibility of cases in which other mitigating factors are present that, only together with determinism, serve to fully exonerate the agent. If cases like this are possible, then the truth of determinism *in conjunction with other mitigating factors* might sometimes suffice to get a person off the hook. Mitigating soft compatibilists who want to make room for such cases must therefore concede that the truth of determinism can be relevant to whether a person is blameworthy for something he did, though, again, they will continue to insist that apart from other mitigating factors—factors that are presumably absent in many (and perhaps most) ordinary situations—determinism does not automatically preclude responsibility.

According to mitigating soft compatibilists, then, the truth of determinism would rarely, if ever, affect whether we can legitimately hold one another morally accountable.

---

[5] In this and the preceding paragraph I have been focusing on mitigating soft compatibilism as it pertains to judgments about blameworthiness and deserved punishment. However, similar remarks apply *mutatis mutandis* when the issue is praiseworthiness and deserved reward.

[6] I thank an anonymous referee for drawing my attention to this issue.

In this respect, mitigating soft compatibilism is similar to standard versions of soft compatibilism and has many of the same attractions. For instance, it is consistent with the resiliency intuition discussed earlier. There are, however, important differences between mitigating soft compatibilism and other brands of compatibilism, and these differences have potential practical implications that warrant attention.

The discovery that determinism is true, mitigating soft compatibilists will say, would have important implications for how we ought to treat one another. It would not affect whether we can legitimately praise and blame, punish and reward, but it would have implications for how much praise and blame and how much punishment and reward people deserve. In the absence of any ordinary mitigating factors, we often tend to assume that wrongdoers bear full (i.e., unmitigated) responsibility for their actions. But if we were to discover that determinism is true, mitigating soft compatibilists, unlike other compatibilists, would insist that this natural tendency would need to be curbed. For instance, they might recommend that we be less resentful and indignant toward others when they do wrong, that we feel less guilty when we do wrong ourselves and that, other things being equal, we punish wrongdoers less severely than we otherwise would have. In their opinion, then, although the discovery that determinism is true would not affect whether we can legitimately continue to hold each other morally accountable and thus would not require us to fundamentally alter our practices of praising, blaming, punishing and rewarding, it would require that we modify these practices in more subtle ways.

An example will help to illustrate the mitigating soft compatibilist's position. Consider a deterministic universe in which Fran, a normal, mentally stable, morally competent adult, murders her sworn enemy, Gina. Fran does this intentionally, we will

suppose, despite knowing that it is morally wrong for her to do it, and her action is deterministically and non-deviantly caused by its immediate psychological antecedents (e.g., a desire to kill Gina and an intention to do so straightaway), each of which can in turn be traced to a deterministic sequence of events that originated in factors beyond Fran's control and for which she bears no moral responsibility. With one important difference, Fran is much like Ann, who was discussed earlier. Recall that Ann shot and killed Bob, despite knowing that it was wrong to do so, and that she satisfied a robust set of compatibilist-friendly conditions for free agency and moral responsibility. Fran too satisfies those same conditions. At the time of the murder, she was not insane, nor was she manipulated or coerced, she did exactly what she wanted to do and, indeed, exactly what she wanted to want to do, she identifies with and wholeheartedly endorses the desires that motivated her murderous behavior, she has and regularly exercises powers of self-control, has the ability to regulate her behavior in accordance with both prudential and moral reasons, and could have done otherwise than kill Gina (on a compatibilist reading of "could have done otherwise"). The main difference between Ann and Fran is that Ann was indeterministically caused to act and, in addition to satisfying numerous compatibilist-friendly conditions for free agency and moral responsibility, satisfied conditions required by various incompatibilist accounts of these things as well, whereas Fran was deterministically caused to act and, consequently, did not satisfy these further incompatibilist requirements.

What would compatibilists say about Fran's case? In particular, what would mitigating soft compatibilists say about it? Earlier I noted that many soft compatibilists would deem Ann morally responsible for killing Bob, and assuming that hers was not a

case of indirect or derivative responsibility (i.e., a case in which the person's moral responsibility for an action depends on or derives from her moral responsibility for something else she did or failed to do), they would deem her directly or non-derivatively morally responsible. On the assumption that acting freely is required for direct moral responsibility, they would also judge that Ann freely killed Bob. Given the similarities between Ann and Fran, most soft compatibilists, including most mitigating soft compatibilists, would make similar judgments about Fran. Because Fran satisfies a robust set of conditions that many contemporary compatibilists regard as sufficient for acting freely and being morally responsible, these compatibilists would say that Fran freely killed Gina, that she is (directly or non-derivatively) blameworthy for killing her and, given the heinous nature of her crime, that she deserves a pretty severe punishment. However, mitigating soft compatibilists will also insist that, because Fran's action resulted from a deterministic causal sequence initiated by factors beyond her control and thus for which she is not morally responsible, she does not bear full or unmitigated responsibility for her action and therefore deserves less blame and a less severe punishment than she otherwise might have if her action had not resulted from a causal sequence of that sort.

Is the mitigating soft compatibilist's verdict about Fran plausible? Why might someone think that being deterministically caused to act by factors beyond one's control mitigates (without necessarily eliminating) one's moral responsibility? In short, why might someone think that mitigating soft compatibilism is true? These are important questions the essence of which will reemerge later in the form of a challenge recently posed to would-be advocates of the view. To set the stage for that challenge and my

suggestions as to how mitigating soft compatibilists might go about addressing it, I turn now to a popular style of argument against compatibilism and a mitigating soft compatibilist response to it.

3. Manipulation and Mitigated Blame

Arguments against compatibilism sometimes feature thought-experiments in which one individual—a deity, devil or sophisticated behavioral engineer, for example— surreptitiously makes another person acquire certain psychological features that in turn deterministically result in that person's performing a certain action. The claim on the part of some incompatibilists is that agents who are manipulated in this way are not to blame for the resulting action, and that, when it comes to blameworthiness, there are no relevant differences between agents in suitably constructed manipulation scenarios of this sort and ordinary agents in deterministic universes. If they are right, then compatibilism is false. The mitigating soft compatibilist, we will see, has a unique response to this sort of argument, one not available to proponents of ordinary brands of compatibilism.

Several manipulation arguments of the sort just described have been offered in recent years.[7] Although the details of these arguments sometimes differ markedly, there nevertheless seems to be a basic strategy that is common to most of them, which can be captured in the following three-step blueprint. *Step one*: tell the story; construct a coherent example in which someone is covertly manipulated into behaving in a certain way. *Step two*: make the case that, when it comes to moral responsibility, there are no important differences between the victim of this sort of manipulation and normal agents

---

[7] See Kane 1996, Mele 2006, Pereboom 2001, Rosen 2002, Taylor 1974 and Todd 2011. For a helpful discussion of the various forms manipulation arguments take, see Mele 2008.

in deterministic universes. *Step three*: make the case that the agent featured in the manipulation story does not act freely and is not morally responsible for what he did. The incompatibilist conclusion follows from steps two and three.

Consider this widely discussed manipulation story of Derk Pereboom's:

Plum is like an ordinary human being, except that he was created by a team of neuroscientists, who…have programmed him to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the result that in the circumstance in which he now finds himself, he is causally determined to undertake the moderately reasons-responsive process [of deliberation] and to possess the set of first- and second-order desires that result in his killing Ms. White. He has the general ability to regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and accordingly he is causally determined to kill White for these reasons. Nevertheless, he does not act because of an irresistible desire. (2001: 113-114)

Now consider the following argument against compatibilism based on this story:

1. When it comes to free action and moral responsibility, there is no relevant difference between Plum and ordinary agents in deterministic universes.

2. Plum did not kill Ms. White freely and, consequently, is not even partly morally responsible for killing her.

3. Therefore, determinism is not compatible with free action and moral responsibility.[8]

Let's briefly consider both of the argument's premises in turn.

Begin with premise 1. Whether this premise is true depends on whether there is a necessary condition for free action and moral responsibility that is not satisfied in Pereboom's story and that could be satisfied by normal agents in deterministic universes. Is there any such condition? Some think not. Pereboom deliberately constructs the example in such a way that Plum satisfies the central conditions for moral responsibility specified by several influential compatibilist accounts of freedom and responsibility. Others, however, demur, insisting that there are plausible compatibilist requirements for free action and moral responsibility that Plum does not satisfy.[9] Suppose theorists in this latter group are right. Then premise 1 is false and the argument is unsound. But might Pereboom's story be augmented so that premise 1 comes out true? I discuss that question shortly. First, some brief remarks about premise 2 are in order.

Why think premise 2 is true? Why think that Plum did not kill Ms. White freely and that he is not morally responsible for killing her? Here an incompatibilist is likely to make a direct appeal to our intuitions. "Look," an incompatibilist might say, "how can we legitimately blame Plum for killing Ms. White or say that he deserves punishment for what he did? It's simply not his fault that he killed her. If anyone is to blame for Ms. White's death, and if anyone deserves to be punished for killing her, surely it's the neuroscientists. True, Plum intentionally murdered poor Ms. White, but he was in effect

---

[8] This argument, it should be noted, is not Pereboom's, though it does has certain affinities with his "Four-case Argument." For the details of that argument, see Pereboom 2001.
[9] See, e.g., Demetriou 2010 and Mele 2006: 138-144.

the puppet of the neuroscientists—a very fancy puppet, to be sure, one with no literal strings and lots of psychological bells and whistles, but a puppet nonetheless. The neuroscientists set in motion processes over which Plum had no control that inevitably resulted in his killing Ms. White. When we keep these things firmly in mind, it is surely implausible to think that Plum acted freely or that he is morally accountable for what he did." Or at any rate, so some incompatibilists will say.

What should compatibilists say in response to the sort of manipulation argument presently at issue? As Michael McKenna has observed, there is no one-size-fits-all compatibilist reply (2008: 143). Depending on the details of a particular compatibilist's preferred account of free action and moral responsibility, it may be that Plum fails to meet a requirement specified by that account. If so, then compatibilists who endorse the account can resist premise 1. This is what McKenna calls a soft-line reply. However, many compatibilists are not content to rely solely on this sort of response.

A difficulty with soft-line replies, as some theorists see it, is that they simply invite the incompatibilist to revise the manipulation story at issue so that the requirement allegedly not satisfied in the original version of the story is satisfied in the revised version. One way to forestall this sort of rejoinder would be to identify a necessary condition for free action and moral responsibility that cannot be satisfied in a manipulation scenario like Pereboom's but that can be satisfied by ordinary agents in a deterministic universe. However, some compatibilists are less than optimistic about the prospects for finding such a requirement and thus are unwilling to rely solely on soft-line replies. In their opinion, these replies may succeed against particular manipulation arguments that appeal to cases that have not been crafted carefully enough so as to ensure

that the agent featured in the example satisfies certain plausible compatibilist requirements for free action and moral responsibility, but in the absence of a requirement of the sort just described, soft-line replies will be fundamentally incapable of rebutting the best, most carefully crafted manipulation arguments, as incompatibilists will be capable of constructing the manipulation story at issue in such a way that all plausible compatibilist-friendly requirements for free action and moral responsibility are satisfied by the featured agent.[10] If this is right, then compatibilists hoping to put manipulation arguments to rest once and for all must take a harder line.

A second option for compatibilists is to grant premise 1, at least for the sake of argument, and to reject premise 2 instead. This is what McKenna calls a hard-line reply. Hard-liners grant that, when it comes to free action and moral responsibility, there is not, or at least needn't be, an important difference between an agent like Plum who has been covertly manipulated and normal agents in deterministic universes. They contend that, because the story about Plum is, or at least could be transformed into, one in which Plum satisfies their preferred set of sufficient conditions for free action and moral responsibility, he did indeed kill Ms. White freely and is morally responsible for killing her, the neuroscientists' involvement notwithstanding.

Proponents of manipulation arguments often attempt to move (via a claim like premise 1) from the intuition that the agent featured in the manipulation scenario does not act freely and is not morally responsible to the conclusion that normal agents in deterministic universes do not act freely and are not morally responsible either. They begin with what they see as the obvious fact that the manipulated agent lacks free will and is not morally responsible and then, on the basis of the likes of premise 1, invite us to

_____

[10] McKenna 2008 is one compatibilist who seems to endorse this take on soft-line replies.

conclude that normal agents in deterministic universes do not act freely and are not morally responsible. One strategy for motivating the hard-line reply is to move in the opposite direction. Consider a version of Pereboom's example *sans* the team of neuroscientists in which determinism is true and Plum is created in the normal way. The hard-liner points out that in this modified version of the story, there is no non question begging reason to suppose that Plum lacks moral responsibility for his action. But now, if there really is no relevant difference with respect to moral responsibility between this version of the case and the original version in which Plum is the victim of the neuroscientists' machinations, then, the hard-liner will insist, there is no non question begging reason to suppose that Plum lacks moral responsibility in the original case.[11]

What are we to make of the hard-line reply? Although some compatibilists apparently feel comfortable saying that an agent like Plum is, or at least could be, morally responsible for his action despite its strange causal history, other theorists, including reflective agnostics about these issues and at least some compatibilists, may find themselves slightly more reluctant to make that concession. They may agree that the incompatibilist reaction to Pereboom's story that I sketched earlier has some intuitive appeal, enough to make them hesitant to grant that Plum acted freely or that he is morally responsible. For theorists of this sort, I recommend considering a mitigating soft compatibilist variation on the hard-line reply.

Mitigating soft compatibilists, of course, will say that upon learning about how Plum's action was produced, we should blame him less harshly and punish him less severely than we would have had his action been produced in a suitably indeterministic fashion. However, on their view, Plum still merits some blame for what he did, assuming

---

[11] Both McKenna 2008 and Fischer 2011 employ this strategy.

that he satisfies their preferred set of compatibilist sufficient conditions for responsibility. According to mitigating soft compatibilists, then, Plum did indeed kill Ms. White freely and is blameworthy for killing her, but his blameworthiness is mitigated by the fact that his action was deterministically caused by events that occurred at or just after his conception over which he had no control and for which he bears no moral responsibility. How plausible is their assessment of Plum?

Imagine that we first learned about the murder of Ms. White in the morning paper. "Extra, extra, read all about it: prominent professor bludgeons homely maid to death with a candlestick in the ballroom!" On first reading about Plum's crime, and in the absence of any obvious mitigating or exculpating circumstances, most of us would no doubt be inclined to blame him fully for what happened. But suppose we subsequently learned about Plum's creation and programming at the hands of the neuroscientists. It is not at all implausible that learning these additional details would mollify, at least to some degree, the anger and indignation we feel towards Plum. Nor is it clear that this shift in our reaction to Plum and his behavior would be unwarranted. At the same time, however, neither is it obvious, at least by compatibilists' lights, that learning about these additional details would completely eliminate our anger and indignation towards Plum or that it would render such attitudes inappropriate given that, the manipulation notwithstanding, Plum retained various morally relevant agential capacities and satisfied a robust set of compatibilist conditions for free agency and moral responsibility.

Mitigating soft compatibilists, it seems, can explain this reaction to the story quite nicely. Again, they will insist that, because Plum satisfies a robust set of compatibilist sufficient conditions for free action and moral responsibility, he did indeed kill Ms.

White freely and is deserving of at least some blame for doing so. However, given the way in which his action was produced, they will also insist that he is worthy of less blame than he would have been had his action been produced in a suitably indeterministic fashion. If they are right, then it would be entirely appropriate for our initial outrage towards Plum to be mollified to some degree upon learning the details of his history, though some anger and indignation would still be warranted given that he satisfied the pertinent compatibilist conditions. Their assessment of Plum thus fits one natural reaction someone with compatibilist sympathies might have to Pereboom's story.

Notice that the mitigating soft compatibilist response to manipulation arguments should still be classified as a hard-line reply insofar as it maintains that Plum did act freely and that he deserves at least some blame for what he did despite his having been manipulated. However, it is a significantly softer hard-line reply insofar as it grants that the strange causal history of Plum's action diminishes his blameworthiness. Unlike other compatibilists who opt for the hard-line response, then, mitigating soft compatibilists are not committed to what some may see as the exceedingly strong claim that learning about the neuroscientists' involvement should occasion no revision whatsoever of our judgments about Plum and his moral responsibility. For this reason, their take on the hard-line reply seems to lighten the compatibilist's burden significantly and thus might make that sort of reply more appealing for those who initially find it difficult to accept.

The preceding discussion suggests a more general point. A standard complaint on the part of some critics of compatibilism is that compatibilist accounts of free action and moral responsibility give various (supposedly wide-spread) incompatibilist intuitions short shrift and that these accounts do not and, indeed, cannot explain away or otherwise

do justice to such intuitions. (Manipulation arguments can be a particularly forceful and evocative way of expressing this sort of complaint.) A virtue of mitigating soft compatibilism is that it seems less vulnerable to this sort of worry. Because the view acknowledge that determinism is incompatible with certain kinds of morally relevant control and that, for this reason, the truth of determinism mitigates moral responsibility, it seems that the view can partially (though perhaps not fully) accommodate many of the intuitions that motivate people to embrace incompatibilism. Consequently, theorists who feel the "pull" of certain incompatibilist arguments and intuitions but who, like Fischer and other soft compatibilists, are hesitant to concede that our status as free agents who are sometimes morally responsible for our behavior hinges on whether determinism is true may find mitigating soft compatibilism to be an especially attractive compromise, as it successfully insulates, to some degree at least, our ordinary beliefs about ourselves as free and responsible agents and our associated practices of holding each other morally accountable from certain empirical discoveries, while at the same time granting that being deterministically caused to act by factors beyond one's control and for which one is not responsible is not entirely irrelevant to an assessment of one's moral responsibility.

4. A Challenge For Mitigating Soft Compatibilism

Earlier I raised the question of whether mitigating soft compatibilism, whatever virtues it may possess, is ultimately a tenable position. I turn now to a recent challenge to mitigating soft compatibilism based on manipulation cases which bears directly on that question, and I sketch a pair answers to the challenge.

Consider Plum once more, and assume for the sake of argument that there is no relevant difference when it comes to free action and moral responsibility between him and ordinary agents in deterministic universes. According to Patrick Todd (2011), standard compatibilists who grant this assumption are committed to what he and most other incompatibilists see as the "excessively strong claim" that "finding out about the role of the neuroscientists should make *no difference* to one's feelings of moral disgust towards Plum….As the incompatibilist sees things," he tells us, "it is hard enough to maintain that Plum deserves blame, but things are apparently even harder: if the [standard] compatibilist is to be believed, not even a revision of judgment is appropriate, given knowledge of Plum's background" (130-131). In Todd's opinion, then, standard varieties of compatibilism come with an even steeper price tag than one might originally have thought. Not only are standard compatibilists committed to saying that Plum is blameworthy for killing Ms. White, they must also make the stronger and seemingly implausible claim that the neuroscientists' involvement in the production of Plum's action in no way affects Plum's moral responsibility. This, Todd thinks (and I agree), seems rather implausible.

Mitigating soft compatibilism, however, apparently does not come with such a high price tag, as proponents of the view are not committed to the excessively strong claim Todd identifies—indeed, they are committed to rejecting it. As I explained in the preceding section, although mitigating soft compatibilists will insist that Plum is morally responsible and that he deserves at least some blame for what he did, they will also claim that his blameworthiness is mitigated and that, upon learning about how Plum's behavior was produced, we should judge him less severely than we otherwise might have had we

20

not been privy to the relevant details of his past. They are therefore in agreement with Todd that finding out about the neuroscientists' involvement should make a difference to our feelings of moral disgust towards Plum and to our assessment of his blameworthiness. It would appear, then, that mitigating soft compatibilism has an important advantage over standard varieties of compatibilism.

Appearances can be deceiving, however. In Todd's opinion, the retreat to mitigating soft compatibilism only gets compatibilists out of the frying pan and into the fire. He says that "if the compatibilist admits that determinism itself is mitigating, a fair question is, In virtue of what? What is it about determinism's obtaining that makes revised judgments of blameworthiness appropriate?" The challenge for mitigating soft compatibilists is to answer this question, and Todd is not sanguine about their prospects for answering it. He claims that the compatibilist who is prepared to agree that determinism mitigates responsibility "is on thin ice, for she must specify features of determinism that only mitigate responsibility rather than ruling it out. Now, what could such features be? I submit that I cannot see what the compatibilist could offer here" (131). In contrast to Todd, I do not find it that difficult to see what mitigating soft compatibilists might say in response to the challenge he poses and thus am slightly more optimistic than he is about their prospects for answering it. In what follows, I explore two ways (there might, of course, be others) in which mitigating soft compatibilists might attempt to answer Todd's challenge.

5. Addressing the Challenge

Determinism, it seems clear, does not preclude all forms of control. This is something that incompatibilists should and, for the most part anyway, typically do concede. When Al drives his car in ordinary conditions, he controls the turns it makes in a way in which his passengers and others do not, even if the universe happens to be deterministic.[12] The debate between compatibilists and incompatibilists is therefore not a debate over whether determinism is compatible with control but rather whether determinism precludes certain kinds of control that we typically believe to be implicated in free action and moral responsibility.

According to Alfred Mele, "A distinction can be drawn between compatibilist or 'proximal' control and a kind of control that requires the falsity of determinism—'ultimate' control" (2006: 7). Indeed, there may be several varieties of both proximal and ultimate control. One way mitigating soft compatibilists might try to respond to Todd's challenge would be to exploit the fact that there are these different kinds of control, some of which are compatible with determinism, others of which are not. One of their basic ideas, again, is that the strongest sort of control minimally required for free action and some degree of moral responsibility is compatible with determinism, but that there are other kinds of control which are required for full or unmitigated responsibility and that are incompatible with determinism. Let's explore one way in which this basic idea might be fleshed out and used in response to Todd's challenge.

Imagine a driving instructor who is happy to let his student, Al, steer the car right, but if Al shows any inclination of steering the car left or of continuing straight, the instructor is determined to take control of the car and steer it right. As it happens, Al steers the car to the right entirely of his own accord. By doing so, he guides the car in a

---

[12] For this example, see Mele 2006: 7.

certain direction, thus controlling the movement of the car, but given the instructor's intention, it seems he lacks control over whether or not the car goes left or right. In this case, the control Al exercises over the movement of the car consists not in his having control over whether the car goes right or left but rather in his guiding the car to the right.

Fischer and his co-author Mark Ravizza (1998: 32) use examples like this to illustrate the difference between two kinds of control: *guidance* control and *regulative* control. Regulative control, they say, implies that the agent has access to alternative possibilities. It "involves a *dual* power: for example, the power to freely do some act *A*, and the power freely to do something else instead," whereas guidance control is a matter of the action issuing from the agent's own suitably reasons-responsive mechanism (1998: 31). Although regulative control and guidance control typically go together, evidently they can be pried apart. In particular, it seems that a person can exercise guidance control even if he lacks regulative control over the event in question. This is illustrated by the likes of the driving instructor example. Al guides the car right and does so, we may suppose, on the basis of his own suitably reasons-responsive mechanism, but owing to the presence, power and intentions of the instructor, he does not have regulative control over the direction the car travels.

Guidance control, Fischer and Ravizza point out, is compatible with the truth of determinism, though regulative control arguably is not. Guidance control is therefore a variety of proximal control, and regulative control, if it is indeed incompatible with determinism, is a variety of ultimate control.[13] They also contend that, whereas guidance control is required for moral responsibility, regulative control is not. Their view is that

---

[13] Mele (2006: 25 n. 4) himself notes that Fischer and Ravizza's distinction between guidance control and regulative control is similar to his distinction between proximal and ultimate control.

regulative control is largely irrelevant to questions about moral responsibility. However, a compatibilist who finds the basic contours of Fischer and Ravizza's brand of (soft) compatibilism appealing needn't share their view about the irrelevance of regulative control to questions of moral responsibility.

It is open to mitigating soft compatibilists to maintain that agents who merely exercise guidance control bear less moral responsibility for their behavior—deserve less blame for their bad actions and less praise for their good actions, for example—than they otherwise would if, in addition to exercising guidance control, they also possessed regulative control over what they did. According to this brand of mitigating soft compatibilism, exercising guidance control over one's conduct is sufficient (in conjunction with any additional compatibilist requirements for free action and moral responsibility) to warrant some praise and blame, punishment and reward. However, agents who also exercise regulative control over their behavior bear greater responsibility for what they do than agents who exercise guidance control alone. This, a mitigating soft compatibilist might say, is because the actions of agents who have regulative control are "up to" those agents in a way that the actions of agents who merely exercise guidance control are not. Agents who have regulative control possess a kind of control over their behavior that allows them to select from two or more alternative courses of action, each of which is genuinely available to the agent, a fact that a mitigating soft compatibilist might plausibly insist is relevant to these agents' moral responsibility.

What would mitigating soft compatibilists who take this line say about the Frankfurt-style cases, which many believe show that regulative control is not required for, and, indeed, is entirely irrelevant to, a person's moral responsibility for what he did?

One option would be to deny that such cases are counterexamples to the claim that regulative control is required for responsibility. I myself find this option quite appealing.[14] Of course, a mitigating soft compatibilist who said this but who rejected the claim at issue would then have to find independent grounds for eschewing that claim. A second option would be to say that, because the agent in a suitably constructed Frankfurt-style case found himself, through no fault of his own, in circumstances which guaranteed that he would perform a certain action, he bears less responsibility for performing that action than he would have if it had been genuinely up to him (in a sense of "up to him" that involves having access to alternative possibilities) whether he performed it.[15]

In a typical Frankfurt-style case, the circumstances that preclude the agent's having regulative control over his action are said to in no way figure in the production of that action, and proponents of Frankfurt-style cases often make the claim, C, that circumstances that are in this way explanatorily irrelevant have no bearing whatsoever on an agent's moral responsibility for what he did. If C were true, then a mitigating soft compatibilist could not plausibly say that the circumstances which preclude the agent's having regulative control in Frankfurt-style cases thereby also diminish the agent's moral responsibility. However, an argument for C is needed. If facts that play no role in producing an agent's action can affect whether the agent possesses regulative control over that action, why can't those same facts affect whether, or to what degree, the agent is morally responsible? Moreover, there is reason to think that C is false.

---

[14] For some recent defenses of this option, see Franklin 2011, Palmer 2011 and Robinson 2012.
[15] Yet another option would involve a combination of the first two responses. A mitigating soft compatibilist might insist that some Frankfurt-style cases are not counterexamples to the claim that regulative control is required for responsibility, while others are.

Marla is walking by a lake when she notices a small child drowning in it. She deliberates for a moment about what to do, comes to the conclusion that she ought (i.e., is morally required) to try to rescue the child and, after a brief moment of hesitation, freely jumps in and rescues him. Compare two different versions of this scenario. In the first version, Marla experiences some slight trepidation about jumping in and attempting the rescue, which she easily overcomes. In the second, Marla suffers from a severe case of hydrophobia, but heroically overcomes her fear. It is plausible that in the first version of the story, Marla merits some praise for her heroic actions. It is also plausible that in the second version she merits even more praise, and her hydrophobia seems relevant here. Had she not been so afraid of water in the second version of the story, she arguably wouldn't deserve any more praise for jumping in and rescuing the child than she deserves in the first version in which she experiences only a minimal amount of fear. But her hydrophobia needn't have played a role in the production of her actions. Marla rescued the child in spite of her fear, not because of it. If this is right, then it would appear that factors that do not figure in the production of a person's action can nevertheless still have an effect on the degree to which the person is morally responsible for performing the action and thus are not always irrelevant to an agent's moral responsibility.[16]

A mitigating soft compatibilist who adopts the view that lacking regulative control, although not sufficient on its own to deprive an agent of responsibility, nevertheless mitigates responsibility has an answer to Todd's challenge. Why does determinism mitigate moral responsibility without eliminating it? One reason, this type of mitigating soft compatibilist will say, is that determinism is incompatible with regulative

---

[16] A similar, though importantly different, point is made by Widerker 2003. Widerker argues that factors that do not figure in the production of an action can affect whether the agent is blameworthy at all.

control (on this, see Fischer and Ravizza 1998: 17-24), and although regulative control is not required for free action and some degree of moral responsibility, agents who have regulative control over their behavior merit more praise and blame for what they do than agents who do not. Agents in deterministic universes thus bear less moral responsibility for their actions than their counterparts in indeterministic universes who have regulative control over at least some of what they do.

Several recent incompatibilist theories of free action and moral responsibility emphasize ultimate responsibility, where *S* is ultimately responsible for *A*-ing only if *A* did not result from a deterministic causal process that was initiated by factors beyond *S*'s control and for which *S* is not responsible. Agents who bear ultimate responsibility for their actions are said to be the ultimate *sources* or *originators* of their behavior in the sense that their actions are not already in the cards and thus cannot be fully accounted for by causal factors outside the wills—beliefs, desires and intentions—of the agents themselves.[17] So, for instance, Robert Kane writes:

> Agents in [suitably indeterministic worlds] have the *power to make choices for which they have ultimate responsibility*. They have the power to be the *ultimate produces of their own ends*…[T]hey have *the power to make choices which can only and finally be explained in terms of their own wills* (i.e., character, motives and efforts of will). No one can have this power in a deterministic world. (1989: 254)

---

[17] Cf. Kane 1989, 1996 and Pereboom 2001.

In the view of some incompatibilists, no agent who lacks the sort of power Kane describes here has free will or is truly morally responsible for his actions.

A second mitigating soft compatibilist reply to Todd's challenge features ultimate responsibility. It is open to proponents of the view to say that in order to act freely and be morally responsible for their actions, agents needn't make choices for which they are ultimately responsible and thus needn't be the ultimate sources or originators of their behavior (at least not in a sense that requires indeterminism), but that agents who are not the ultimate sources of their behavior bear less moral responsibility for their actions than those who are. According to this brand of mitigating soft compatibilism, agents who are the ultimate sources of their behavior are worthy of more blame and deserve more severe punishment for their bad actions, and are worthy of more praise and deserve greater reward for their good actions, than agents who lack ultimate responsibility, as their actions can, in Kane's words, "only and finally be explained in terms of their own wills." And, because agents in deterministic universes lack the power to make choices of which they are the ultimate sources and for which they are ultimately responsible, it follows that they bear less moral responsibility for their actions than agents in indeterministic universes who have and exercise that power.

Consider an analogy. We might liken the world to a book and each fact about the world a sentence in that book. When agents perform actions, it is as if they are writing a new sentence in the book of the world. Now, in deterministic universes there is a sense in which all the sentences in the book of the world have already been written, since they are all entailed by the conjunction of the laws of nature (which, of course, are themselves sentences in the book) and sentences written at the very beginning of the book describing

the initial state of the universe. But if determinism is false and agents can sometimes be the ultimate sources of their choices and subsequent actions, they sometimes write sentences in the book of the world that are not entailed by previous sentences and that can only and finally be traced back to the wills of the agents themselves. In this way, agents who are ultimately responsible sometimes make genuinely novel contributions to the world, contributions that were not already in the cards, as it were, whereas agents in deterministic universes do not make any such contribution, since any contribution they make was implicit in what was already written. A mitigating soft compatibilist might plausibly claim that agents whose actions do not constitute truly novel contributions to the book of the world bear less responsibility for what they do than agents whose actions do constitute such a contribution and that, for this reason, although agents in deterministic universes can act freely and be morally responsible for their actions, they deserve less praise or blame, and less punishment or reward, than their counterparts in indeterministic universes who are the ultimate sources of at least some of their behavior.

Todd challenges mitigating soft compatibilists to explain why the truth of determinism mitigates blameworthiness without eliminating it. I have sketched two ways in which a mitigating soft compatibilist might attempt to meet this challenge. According to the first, determinism is incompatible with agents possessing regulative control over their behavior, and although agents who lack regulative control may still act freely and be morally responsible, they bear less responsibility for their actions than their counterparts in indeterministic universes who possess regulative control over at least some of what they do. According to the second, determinism is incompatible with agents bearing ultimate responsibility for, and thus with their being the ultimate sources of, their actions,

29

and although agents who are not ultimately responsible for their behavior may still act freely and be morally responsible, they bear less responsibility for their actions than their counterparts in indeterministic universes who are the ultimate sources of at least some of what they do. A comprehensive defense of mitigating soft compatibilism would require filling in the details of these responses (for example, by offering further support for the claim that lacking regulative control over, or lacking ultimate responsibility for, one's actions mitigates without necessarily eliminating praiseworthiness and blameworthiness), but that is a project for another occasion. My aim here has been to show that this particular brand of compatibilism is far from indefensible and to highlight ways in which proponents of the view might try to defend it. However, one worry about the responses I have suggested on behalf of mitigating soft compatibilism should be addressed now.[18]

Neither regulative control nor ultimate responsibility, it might be argued, admits of degrees. This seems especially clear in the latter case. It is difficult to see how one person who is the ultimate source of his actions could be any more or less ultimately responsible for what he does than another person who is also the ultimate source of her actions. But if having regulative control over, and being ultimately responsible for, one's behavior are not matters of degree, this might seem to generate a puzzle. How, a critic might wonder, can something that does not itself come in degrees nevertheless confer a greater degree of moral responsibility?

Ultimate responsibility, I am willing to concede, is not a degreed notion, but what about regulative control? It is widely agreed that control comes in degrees. Like most ordinary people, I don't have any special powers over dice, but I might nevertheless have some control over whether I roll a six with a fair die, if, for example, I have control over

---

[18] I thank an anonymous referee for urging me to address the worry.

whether I roll at all. Someone who does have special powers over dice, however, might have more control over whether he rolls a six than I have over whether I roll one. Now, if control comes in degrees, why not *regulative* control specifically?

Suppose a wealthy gambler offers both Dan and me (who, being philosophers, are people of extremely modest means) the following bet: "If you both roll a six, I'll give you each a million dollars; but if either of you fail to roll a six, then you must each pay me a million dollars." Unsurprisingly, Dan and I decline to accept the wager and, accordingly, refrain from rolling. But could we each have rolled a six? Yes; it was within my power, let's suppose, to roll (i.e., I had regulative control over whether I rolled), and had I rolled, I might very well have rolled a six. It was also within Dan's power to roll, and every time he rolls there is a ninety percent chance that he will roll a six if he tries to roll one, as Dan has special powers over dice that I lack. Does Dan have more regulative control over whether he rolls a six than I have over whether I roll one? It certainly sounds like it.

The preceding remarks are suggestive, but they are hardly decisive. So let's proceed on the assumption that neither regulative control nor ultimate responsibility come in degrees. This assumption by itself poses a problem for my suggestions as to how mitigating soft compatibilists might respond to Todd's challenge only if the following thesis is true: a factor F confers an additional degree of responsibility on an agent only if F itself is something that admits of degrees. However, this thesis is false.

It is widely recognized that a person needn't have acted intentionally in order to be responsible for what he did.[19] I might be to blame for stepping on your toe even if I didn't step on it intentionally—perhaps I was being careless and wasn't watching where I was going. Nevertheless, whether an action is intentional arguably can affect the amount

---

[19] See Mele and Sverdlik 1996.

of blame its agent deserves for performing it. If I were to intentionally step on your toe (in an effort to cause you pain, say), it seems that I would deserve even more blame than I would have had I stepped on it unintentionally but negligently. Notice, however, that an action's being intentional is something that does not seem to admit of degrees. Evidently, then, an action's being intentional is a factor that can confer an additional degree of blameworthiness even though it is not something that comes in degrees. And if that's right, then it's hard to see how the fact (if it is indeed a fact) that neither regulative control nor ultimate responsibility admit of degrees is itself a barrier to their conferring a greater degree of blameworthiness on an agent who possesses them.

6. Conclusion

Compatibilism, again, is the thesis that determinism is consistent with free action and moral responsibility. Compatibilism per se leaves open a variety of positions concerning what implications, if any, the truth or falsity of determinism might have for moral responsibility. I have identified one such position—mitigating soft compatibilism—and have argued that it is far from indefensible and that it has unique resources for dealing with manipulation arguments against compatibilism. Whether this brand of compatibilism is ultimately viable or not will depend, among other things, on whether compatibilism itself is defensible. Insofar as compatibilism is defensible, however, I submit that mitigating soft compatibilism merits further attention by compatibilists and their critics.[20]

References

Ayer, A. J. 1954. Freedom and Necessity, in A.J. Ayer, *Philosophical Essays*. New York: St. Martin's Press: 3-20.

Clarke, R. 2003. *Libertarian Accounts of Free Will*. New York: Oxford University Press.

Demetriou, K. 2010. "The Soft-line Solution to Pereboom's Four-Case Argument," *Australasian Journal of Philosophy* 88: 595-617.

Fischer, J.M. 2006. *My Way: Essays on Moral Responsibility*, New York: Oxford.

Fischer, J.M. 2011. "The Zygote Argument Remixed," *Analysis* 71: 267-272.

Fischer, J.M. and Ravizza, M. 1998. *Responsibility and Control: An Essay on Moral Responsibility*, Cambridge: Cambridge University Press.

Franklin, C.E. 2011. "Neo-Frankfurtians and Buffer Cases: The New Challenge to the Principle of Alternative Possibilities," *Philosophical Studies* 152:2, (2011): 189–207.

Hobart, R.E. 1934. "Free Will as Involving Determinism and Inconceivable Without It," *Mind* 43: 1-27.

Kane, R. 1989. "Two Kinds of Incompatibilism," *Philosophy and Phenomenological Research* 50: 219-254.

Kane, R. 1998. *The Significance of Free Will*, New York: Oxford University Press.

McKenna, M. 2008. "A Hard-line Reply to Pereboom's Four-Case Manipulation Argument," *Philosophy and Phenomenological Research* 77/1: 142-159.

Mele, A. 2006. *Free Will and Luck*, New York: Oxford University Press.

Mele, A. 2008. "Manipulation, Compatibilism and Moral Responsibility," *Journal of Ethics* 12/3-4: 263 - 286.

Mele, A. and Sverdlik, S. 1996. "Intention, Intentional Action and Moral Responsibility,"

Philosophical Studies 82: 265-287.

Nowell-Smith, P.H. 1948. "Free Will and Moral Responsibility," *Mind* 57: 45-61.

Palmer, D. 2011. "Pereboom on the Frankfurt Cases," *Philosophical Studies* 153: 261-272.

Pereboom, D. 2001. *Living Without Free Will*, Cambridge: Cambridge University Press.

Robinson, M. 2012. "Modified Frankfurt-type Examples and the Flickers of Freedom," *Philosophical Studies* 157 :177-194.

Rosen, G. 2002. "The Case for Incompatibilism," *Philosophy and Phenomenological Research* 64: 699-706.

Schlick, M. 1939. "When is a Man Responsible?" in Moritz Schlick, *Problems in Ethics*. New York: Prentice-Hall: 141-158.

Smart, J.J.C. 1961. "Free-Will, Praise and Blame," *Mind* 70: 291-306.

Speak, D. 2008. "Guest Editor's Introduction: Leading the Way," *Journal of Ethics* 12/2: 123-128.

Taylor, R. 1974. *Metaphysics*, Englewood Cliffs: Prentice Hall.

Todd, P. 2011. "A New Approach to Manipulation Arguments," *Philosophical Studies* 152: 127-133.

Widerker, D. 2003. "Blameworthiness, and Frankfurt's Argument Against the Principle of Alternative Possibilities," in Widerker and McKenna (eds.), *Moral Responsibility and Alternative Possibilities*: *Essays on the Importance of Alternative Possibilities*. Aldershot, UK: Ashgate Press: 53-74.