1. INTRODUCTION

Compatibilism is the thesis that the truth of causal determinism would not, all by itself, rule out the possibility of moral responsibility. This view is attractive for several reasons. For example, it seems to appreciate our status as agents in a natural order, an order that very well may be deterministic. Moreover, compatibilism also respects our status as persons. If compatibilism is true, then our morally responsible agency does not, as John Martin Fischer puts it, “hang on a thread,” i.e., it is resilient with respect to the deliverances of contemporary physicists.¹ And an important normative consequence of compatibilism seems to be that even if causal determinism turns out to be true, our practices of blaming and of holding others morally responsible for their wrongdoing might still be justified.

But despite these considerable attractions, compatibilism is not without its detractors.² And over the past two decades, a new anti-compatibilist argument has emerged as a serious threat to compatibilist accounts of moral responsibility. As it is typically presented, this argument—“the manipulation argument”—is driven by the intuition that manipulation of a certain sort undermines agents’ status as morally responsible. And according to the incompatibilist, the best explanation for their non-responsibility in such cases is that the manipulated agent isn’t the causal-historical source of her action. Incompatibilists then point out that in this respect, manipulation looks mighty similar to causal determinism. After all, if causal determinism is true, there are causal antecedents in the distant past that together with the laws of nature entail that I will act in precisely the way I do act. And since I’m not the source of these causal antecedents or the laws of nature, the thought goes, neither am I the source of my actions. Thus, given the apparent structural similarity between manipulation (of a particular sort) and causal determinism, we should conclude that just as manipulation exculpates an agent’s responsibility, so too does causal determinism.³

¹ Cf. Fischer [2006].
² Cf. van Inwagen [1983]; Kane [1996]; Pereboom [2001]; Mele [2006].
³ For the best versions of the manipulation argument, see Pereboom [1995, 2001, forthcoming] and Mele [2006]. Although note that if Patrick Todd [2012] is correct, then we are supposed to directly “see” the truth of incompatibilism itself from the exculpating features of manipulation. As Todd puts it, “it seems to me that the [manipulation argument]
And if that’s not bad enough (for the compatibilist), a further consequence of this argument is that if determinism obtains in the actual world, then none of our practices of blaming or holding others morally responsible are justified. After all, in such a case, no one would be morally responsible for any of their actions, and as a result, no one would be a fitting target of blame. Nor would anyone be an apt candidate for being held responsible for those actions. Moreover, we might worry along with Peter Strawson that if we are not responsible, the very fabric of reciprocal interpersonal relationships would be undermined. And if this is correct, things certainly do look bleak for the compatibilist.

Yet despite this bad prognosis, this anti-compatibilist conclusion strikes me as plausible only because the premises of the manipulation argument rely on problematic assumptions concerning the nature of moral responsibility and blameworthiness. To defend this claim, I introduce the notion of minimal moral responsibility in §2. I then argue in §3 that manipulation as such doesn’t undermine minimal moral responsibility because manipulation doesn’t undermine the fittingness of the reactive emotions of resentment, indignation, and guilt. Similarly, I argue in §4 that manipulation doesn’t undermine minimal moral responsibility because it doesn’t undermine an interpersonally significant form of blame that is particular to our participation in relations of mutual regard. Thus, because manipulation is compatible with two forms of blame that, when fitting, render agents minimally morally responsible, the manipulation argument doesn’t undermine compatibilism.

2. MINIMAL MORAL RESPONSIBILITY

To begin, distinguish between ultimate moral responsibility and minimal moral responsibility. We can say that an agent is ultimately morally responsible for her action(s) only if certain severe, formal forms of moral sanction are fitting—e.g., extended punishment in prison, execution, or perhaps an eternity in hell. In other words, the property of being ultimately morally responsible is essentially tied to

---

4 Peter Strawson [1974]. John Martin Fischer [2006] and Fischer and Ravizza [1998] share a similar worry—that if incompatibilism is true, then the truth of determinism might undermine our status as persons.

5 Throughout the remainder of the paper, for simplicity’s sake, I will refer to “minimal moral responsibility” as “moral responsibility” (unless otherwise specified).

6 In fact, Galen Strawson understands the concept of moral responsibility to be tied to notions of heaven and hell. According to G. Strawson, to be morally responsible for some action is for it to be fitting that God reward you with an eternity of bliss in heaven (if that action is good) or for it to be fitting that God punish you with an eternity of torment in hell (if that action is bad). For more, see G. Strawson [2003].
severe, asymmetric forms of sanction and punishment, which might be meted out by the state or by God. By contrast, an agent is minimally morally responsible for her action(s) only if certain informal forms of moral criticism and sanction are fitting—e.g., expressions of resentment, blame, rebuke, etc. Thus, the property of being minimally morally responsible is essentially tied to blame as it arises in the context of ordinary (reciprocal) interpersonal relationships.\(^7\)

Consequently, both ultimate moral responsibility and minimal moral responsibility are basic desert-entailing conceptions of responsibility.\(^8\) For if \(A\) is ultimately responsible for \(x\)-ing, then \(A\) deserves, in a basic sense of desert that is not reducible to consequentialist or contractualist considerations, to be held ultimately morally responsible for \(x\)-ing. Likewise, if \(A\) is minimally morally responsible for \(x\)-ing, then \(A\) deserves, in a basic sense, to be held minimally morally responsible for \(x\)-ing. Therefore, both ultimate moral responsibility and minimal moral responsibility are features of what Gary Watson has called the “accountability face” of moral responsibility [Watson 1996], which is characterized as the face of responsibility tied to the varied set of morally and legally significant responses to others’ behavior (e.g., blame, sanctions, punishment, etc.).

Now it might be that minimal moral responsibility and ultimate moral responsibility are coextensive, such that whenever an agent is minimally morally responsible, she is also ultimately

---

\(^7\) Thus, as I understand them, theories of responsibility which take Peter Strawson’s “Freedom and Resentment” [1974] as their starting point are theories of minimal moral responsibility because they understand what it is for an agent to be morally responsible in terms of the reactive emotions, which are the “non-detached attitudes” essential to ordinary interpersonal relationships. It seems to me that this is one of Strawson’s most important insights regarding the nature of moral responsibility. And it’s one that was largely ignored prior to Strawson and only haltingly heeded since.\(^7\) To see this, compare Strawson’s conception of moral responsibility as minimal moral responsibility with those of Nietzsche and of Galen Strawson. For Nietzsche, the \textit{causa sui}, i.e., freedom of the will in the “metaphysical superlative sense,” and the kind of moral responsibility that such freedom underwrites, is grounded in “the desire to bear the whole and sole responsibility for one’s actions and to absolve God, the world, ancestors, chance, and society from responsibility for them,” [BGE 21]. This conception of freedom and responsibility is certainly a far cry from Strawson’s idea that morally responsible agency is the glue that holds together ordinary human relationships. Similarly, Galen Strawson, writing two and a half decades after his father, identifies the property of being morally responsible with the property of being an apt target of divine damnation (or beatification, as the case may be). In other words, Galen Strawson simply identifies moral responsibility with ultimate moral responsibility. This, too, takes us away from (Peter) Strawson’s more minimal, but no less robust, conception of moral responsibility as being grounded in participation in distinctively human forms of engagement. For even if Galen Strawson is correct to think that it is impossible for us to be ultimately morally responsible in the sense that it is fitting for God to eternally damn us, why should it thereby follow that we are not morally responsible in the sense that it is fitting for others to engage with us via the reactive emotions? Moreover, other than metaphysical megalomania, why should we even care about a sense of moral responsibility that abstracts away from our ordinary interpersonal relationships?

\(^8\) Since minimal moral responsibility is a form of basic desert-entailing responsibility, which is the notion of responsibility at the heart of the debate (for more on this point, see Pereboom [2001]), I am not guilty of changing subjects by focusing on minimal moral responsibility in this paper.
morally responsible. In failing to distinguish between ultimate and minimal moral responsibility, the moral responsibility literature has proceeded as if this were the case. But it certainly isn’t obvious that whenever an agent satisfies the conditions for minimal moral responsibility, she also satisfies the conditions for ultimate moral responsibility. For it certainly seems plausible to think that given the severe sanctions that I might be met with if I am ultimately morally responsible for my actions, there are higher standards on instantiating the property being ultimately morally responsible than on the property being minimally morally responsible.

And although I admit that perhaps an agent is not ultimately morally responsible if she is subjected to manipulation of a certain sort, the same cannot be said of minimal moral responsibility. As I understand it, our standing as minimally morally responsible agents grounds our participation in meaningful reciprocal interpersonal relationships (most notably e.g., relations of mutual regard). And plausibly, neither manipulation nor the truth of causal determinism entail that we cannot enjoy meaningful reciprocal interpersonal relationships. So, equally plausibly, neither manipulation nor the truth of causal determinism entail that we are not minimally morally responsible. Thus, if we take seriously the idea that minimal moral responsibility is itself a real form of moral responsibility, we will conclude that compatibilism—at least a form of compatibilism according to which causal determinism doesn’t rule out the possibility of minimal moral responsibility (and its attendant blaming practices)—is not threatened by the manipulation argument.9

3. Manipulation and the Reactive Emotions
To defend compatibilism, then, against manipulation worries, I will attempt to show that the first premise of the manipulation argument—the idea that manipulation exculpates responsibility—provides a less powerful reason for doubting compatibilism than is often supposed. In other words, I’m proposing a compatibilist strategy for responding to manipulation arguments which entails that, even in some cases of manipulation, agents might deserve to be blamed for their actions. But compatibilists who, like me, reach this conclusion are accused of the unsavory task of “biting the bullet” with respect to manipulation. Accordingly, this reply is known as hard compatibilism.10 Hard

---

9 And I think it’s only fair that we take minimal moral responsibility seriously since some version of it has been defended by no less than P.F. Strawson [1974], Watson [1987], Wallace [1994], Fischer and Ravizza [1998], Stephen Darwall [2006], and Michael McKenna [2012].

10 Cf. Watson [1999].
compatibilists, then, claim that not only is moral responsibility compatible with causal determinism, but that it is compatible with certain forms of manipulation.\textsuperscript{11}

In defense of hard compatibilism, I will argue that manipulation does not affect the propriety of several interpersonally significant responses to wrongdoing. These responses, which can be accurately described as forms of blame, are appropriate given the interpersonal expressive significance of the wrongs in question. Thus I conclude that manipulated agents can be blameworthy for their action. And if this is correct, then hard compatibilism—like hard liquor or the hard rock stylings of MC5 or Black Sabbath—may not be so hard to stomach after all.

To begin, consider the following instance of manipulation (due to Derk Pereboom) that is thought to undermine the manipulated agent’s responsibility and blameworthiness.

Plum is like an ordinary human being, except that he was created by neuroscientists, who, although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the result that in the circumstances in which he now finds himself, he is causally determined to undertake the moderately reasons-responsive process and to possess the set of first- and second-order desires that results 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 for these reasons. Nevertheless, he does not act because of some irresistible desire [Pereboom 2001, 113-14].\textsuperscript{12}

In this case, Professor Plum murders Ms. White. Yet \textit{intuitively}, says Pereboom, he is not morally responsible for the murder. Neither is he intuitively blameworthy for his crime.

Now, according to Pereboom, the best explanation for why Plum is not responsible is that he is not the ultimate source of his action. But though Pereboom invokes the notion of ultimacy here, he means for this argument to show not only that Plum is not ultimately morally responsible,

\textsuperscript{11} For alternative defenses of hard compatibilism, see Michael McKenna [2008] and John Martin Fischer [2011].

\textsuperscript{12} For those of you keeping score at home, this is Case 2 of Pereboom’s ingenious “Four-Case Argument” against compatibilism. And by only focusing on this case, I’m presenting Pereboom’s case of Professor Plum in much the same way in which Mele [2006] presents the case of Ernie, the person who grows from a zygote created by the goddess Diana. Thus, as I present it here, it is a kind of “Original Design” argument. For a particularly compelling reply specifically to Pereboom’s “Four-Case Argument,” see McKenna [2008].
but that he is not morally responsible in any basic desert-entailing sense, which includes minimal moral responsibility. After all, as I define it, minimal moral responsibility is to be understood (in part) in terms of reactive emotions like resentment and indignation. And these attitudes are widely thought to be basic desert-entailing emotions, such that whenever it is fitting for A to resent B, B deserves to be resented and B deserves the loss of relational goods like continued friendship, presumptions of good will, and trust. Thus, if Pereboom’s argument is successful, then compatibilist theories of minimal moral responsibility are in trouble.

At first blush, then, Pereboom’s case poses significant challenges to the hard compatibilist—even hard compatibilists who are careful to distinguish minimal from ultimate moral responsibility. Therefore, because the hard compatibilist (of the sort I’m envisioning) claims that minimal moral responsibility and its attendant forms of blame are compatible with certain forms of manipulation (e.g., the forms of manipulation present in the case above), she must compellingly address cases like this. That is, she must render it plausible that in this case, Plum is, pace Pereboom, minimally morally responsible and blameworthy in the sense that he is an apt target for morally significant blaming responses.

3.1
To motivate the hard compatibilist’s claim that Plum might be a fitting target of some forms of blame (and thereby, a minimally morally responsible agent), let’s first consider how Ms. White’s family and friends should react to Plum when they discover that he has murdered their loved one.

- Should they be wholly nonplussed about the death of their loved one?
- Should they be sad and disappointed at what Plum has done?
- Should they blame Plum?
- Should hold him responsible in overt ways (through e.g., expressed resentment, rebuke, chastisement, etc.)?

---

13 Strawson [1974] suggests that resentment presupposes that the target deserves to suffer, and there’s a sense in which this is correct, since losing friendships, presumptions of good will, and trust can cause a great deal of suffering. Of course, resentment doesn’t presuppose that wrongdoers deserve to suffer for suffering’s sake. In other words, resentment might be punitive, but we need not understand it as sadistic.
Obviously, without the assumption of manipulation, it would seem appropriate for White’s family and friends to not only feel sad, disappointed, and frustrated at Plum’s wrongdoing, but also to feel resentment towards Plum and to express their blame in overt ways.

But how should their response(s) change if they were to discover that Plum had been causally determined to murder White via covert manipulated by a team of neuroscientists? On the basis of this information:

- Should they mitigate their blame upon learning of Plum’s victimization at the hands of the neuroscientists?
- Should they revise their emotional responses from punitively-tinged states (e.g., resentment, indignation, etc.) to mere sadness or despondence?
- Should they fully withdraw their resentment, indignation, and anger?
- Should they stoically ignore Plum’s agential contributions to their loved one’s death because those contributions can be traced back to causal factors beyond Plum’s control?
- Should they leave their responses wholly intact?

These are difficult questions, and I should say, I do not mean them to be merely rhetorical, as if it were obvious that the right answers to such questions entailed hard compatibilism. I ask them because I think that by sifting through the answers to these questions, we will gain some insight into whether hard compatibilism is a viable position. Of course, it is my view that in answering these questions, we’ll come to see that some forms of blame are appropriately directed at wrongdoers—even if they are the victims of manipulation. And this would entail that some forms of moral responsibility—specifically minimal moral responsibility—are compatible with manipulation (and also, with causal determinism).

But I certainly don’t think that any of this is obvious, and since even good rhetorical questions are no substitute for good arguments, I turn to arguments below.

3.2
First, it would undoubtedly be natural for White’s family and friends to feel resentment and indignation towards Plum upon hearing that Plum had murdered White. But even the incompatibilist
can admit that this response (or at least, the dispositions that lead us to take such a response) is *natural*. The important question, however, is whether, given Plum’s manipulation, these blaming attitudes are *fitting*—i.e., whether Plum *deserves* to be targeted with these attitudes. But if we’re going to figure out whether resentment and indignation are fitting attitudes to Plum’s wrongdoing, we must first figure out what it is to resent (or to be indignant). That is, we must have an account of these blaming attitudes and of their fittingness conditions.

Concerning these attitudes, I conjecture that like other mental states (e.g., beliefs, desires, etc.), the emotions of resentment and indignation have intentional object (i.e., they are *about* something). But notice, if I’m at the bar and a friend offers me a cigarette, I can *believe* that she offers me a cigarette; or if she hasn’t yet offered me one, I can *desire* that she offers me a cigarette; or if I’ve wanted one for a long time and she then offers me one, I can *be happy* that she offers me a cigarette; or if she knows that I’m trying to quit when she offers me a cigarette, I can *resent* that she offers me a cigarette. In each of these cases, the mental states of belief, desire, happiness, and resentment each have an identical intentional object—i.e., a state of affairs that includes the action of my friend offering me a cigarette. Since mental attitudes of different kinds can share intentional objects, it thus seems that the intentional objects of mental states cannot serve to individuate them. So the fact that White’s family and friends resent Plum’s actions doesn’t yet distinguish resentment from other states in their mental economies (e.g., their belief that Plum murdered White, their sadness that Plum murdered White, etc.).

Thus, I propose that the attitudes of resentment and indignation can be distinguished from other mental states by their distinctive *representational qualities*. That is, in resentment, I represent an agent’s action as being *disrespectful* or as showing ill will or disregard to me. And because indignation is plausibly thought of as resentment’s vicarious analog, in feeling indignation, I represent an agent’s action as being disrespectful or as showing ill will or disregard to another participant in the moral community. Thus, in feeling indignation towards Plum, White’s loved ones represent his murdering of White as having wantonly disrespected White’s standing as a

---

14 Following Pereboom, I am taking the relevant notion of desert to be basic in the sense that it is not reducible to consequentialist or contractualist considerations. However, I also accept the following claim: If B is a fitting target of A’s resentment, then B deserves to be resented. Though this claim is not an analysis of desert (or of fittingness), it provides us with some insight into the relationship between these two normative notions.

15 Although I mainly use “disrespect,” I mean for it to be synonymous with “ill will” and “disregard.”

16 For more on a conception of indignation as a vicarious analog of resentment, see Strawson [1974] and Wallace [1994]. I discuss one form of guilt—plausibly thought of as the self-regarding analog of resentment—in §2.5.
participant in the moral community—a person imbued with dignity, who is owed a certain degree of proper respect.17

Accordingly, because resentment and indignation represent others as having disrespected our standing as participants in the moral community, there is some reason to think that the indignation that White’s family and friends feel is fitting in this case. After all, generally speaking, emotions are fitting just in case their intentional content and representational quality fit the world. And in this case, the intentional content—Plum’s murdering of White—obtains. So the real question is whether White’s family and friends accurately represent this content as disrespectful. If Plum’s actions are genuinely disrespectful (despite the manipulation), then resentment and indignation are fitting responses to Plum’s wrongdoing and hard compatibilism is vindicated. But if not, then incompatibilists like Pereboom are correct.

Of course, to defend the hard compatibilist contention that Plum’s actions are actually disrespectful in the relevant way, I must say more about what constitutes the interpersonally significant form of disrespect that is at stake. For it might be that being a victim of manipulation undermines Plum’s status as an agent capable of disrespect. And if so, then the indignation that White’s family and friends naturally feel is, while perhaps excusable, inappropriate.

3.3.
To begin, it’s plausible to think that, following Adam Smith, an agent’s action is disrespectful (in the sense of disrespect at stake in resentment) just in case that action reveals “the little account he seems to make of us, the unreasonable preference which he gives to himself above us, and that absurd self-love, by which he seems to imagine, that other people may be sacrificed at any time, to his conveniency or his humour,” [Smith TMS II.iii.1.5]. Relatedly, we can say, with Kant, that an agent’s action is disrespectful just in case that action uses Humanity merely as a means, or otherwise fails to use Humanity as an end in itself, [Kant G 4:429]. Or finally, echoing T. M. Scanlon, we might say that an agent’s action is disrespectful when it issues from a principle that others could reasonably reject, [Scanlon 1998]. Taken together, Smith, Kant, and Scanlon provide a plausible model for

17 Obviously, the representational quality of an emotion doesn’t itself have to be consciously represented as part of the intentional object of the emotion. Thus, it’s not a feature of resentment that part of the intention object of resentment is a propositional attitude that has “S’s φ-ing was disrespectful.” Instead, resentment itself represents S’s φ-ing as disrespectful.
understanding disrespect of the sort at stake in resentment and indignation, which I will adopt for the purposes of this paper.

Now, as Pereboom tells it, though Plum is manipulated by a team of neuroscientists, he (i) blew off the decisive moral reasons for not killing White, (ii) acted on egoistic considerations, (iii) displayed “that absurd self-love,” and (iv) systematically underestimated the moral significance of White’s standing as an equal person. It seems to me then, that Professor Plum has, in fact, disrespected Ms. White and has offended against her dignity and moral significance. After all, his action conveys a woeful underestimation of the moral status of Ms. White. Moreover, in murdering White for egoistic reasons, he regards her as a mere means to his own self-interested ends, and he fails to appreciate the significance of her consent, since he has acted for reasons that she certainly would have rejected as having any normative significance. It would thus seem that if Smith, Kant, and Scanlon have correctly interpreted what it is to disrespect another agent, then Plum is guilty of disrespecting White.

“But what about the manipulation?” you might ask. Wouldn’t the presence of the manipulators entail that Plum hasn’t really been disrespectful?

In response to this line of questions, I would first note that the fact that a team of neuroscientists has manipulated Plum by causally determining him to deliberate and act in this manner doesn’t uncontroversially vitiate the general claim that deliberating and acting in precisely this manner is what constitutes disrespect of the sort that is represented in resentment and indignation. But furthermore, I think it is actually the case that even though he has been causally determined by a team of neuroscientists, Plum’s action does convey the interpersonally significant form of disrespect. Of course, rather than simply appealing to my own intuitions, I want to offer three independent arguments for thinking that, despite being the victim of manipulation, in killing her for terrible reasons, Plum disrespects White, and is thereby an apt target of resentment and indignation.

A. First, following Peter Strawson, we can recognize that the truth of causal determinism doesn’t entail universal good-will or, to put it in the idiom I favor, universal respect. Consequently, even if determinism were true, we could nevertheless distinguish between good- and ill-will, respect and disrespect, reasonable regard and disregard. And note: this point holds even if the incompatibilist is

18 Strawson [1974].
correct in claiming that determinism, if true, would rule out the possibility of agents being the ultimate causal source of their actions. But this is equally true of manipulation as well. The fact that he was the victim of manipulation surely doesn’t entail that Plum is necessarily respectful, or even merely indifferent towards others. Therefore, just as the truth of determinism doesn’t entail a breakdown in the distinction between respect and disrespect, neither should manipulation. As a result, we have a reason to think that manipulation itself does not undermine an agent’s ability to be disrespectful. And so, we have a reason to think that even though he is manipulated, Plum is disrespectful in his treatment of White.

**B.** The second argument for the claim that Plum’s action was disrespectful begins by acknowledging the agential capacities that are implicated in actions that can be properly characterized as respectful or disrespectful. For example, we don’t think that when animals injure us, or when a baby hurts our feelings (say, by not smiling at us) that they have disrespected us. This is because in order to disrespect someone, agents need to be able to appreciate the normative significance of what Peter Strawson called “the basic demand” for reasonable regard: the demand that we relate to others in ways that respect their status as persons. And of course, by Pereboom’s stipulation, Plum is able to appreciate the reasons that favor not killing White, and he has the general capacity to govern his behavior in light of those reasons. But since he doesn’t care about them—since egoistic reasons are more compelling to him—he acts contrary to moral reasons. And it seems to me that because he recognizes the reasons in question to be binding and yet chooses to act contrary to those reasons, Plum’s killing of White is disrespectful in an especially flagrant way. It seems then, that it is whether an agent is reasons-responsive that determines whether her actions are potentially respectful or disrespectful. And Plum is reasons-responsive. So plausibly, when Plum murders White, he is failing to respect her in interpersonally significant ways.

**C.** Lastly, I want to offer a third argument for thinking that the neuroscientists’ presence in the etiology of Plum’s action is consistent with Plum showing disrespect or ill will in his action. To begin, distinguish between Professor Plum on the one hand and Professor Chum on the other. Like

---

19 For a compelling account of reasons-responsiveness, see John Martin Fischer and Mark Ravizza [1998]. See also, R. Jay Wallace [1994].
Plum, Chum is also the victim of manipulation. But unlike Plum, Chum has been manipulated to save Ms. Knight’s life even though, on the basis of self-interest alone, Chum has most reason to go ahead and let Knight die. Moreover, whereas manipulation causes Plum’s standing intentions and attitudes towards White to seemingly manifest ill-will and disrespect, it causes Chum’s standing intentions and attitudes towards Knight to seemingly manifest good will and respect. For example, Chum correctly regards Knight as having a significant moral value—a value that cannot be overridden by his trivial concerns. And he is suitably sensitive to the fact that her concerns and interests are reasons-giving for him, at least in the sense that they put significant side-constraints on how he can treat her. In light of Chum’s sensitivity to Knight’s concerns, it seems true to say that Chum respects Knight when he consciously decides to save her life on the basis of these reasons.

No doubt, the attitudes, deliberative tendencies, values, motives, intentions, and actions that constitute Chum’s respect of Knight are causally determined according to the manipulators’ plan. But rather than showing that Chum isn’t respectful, it simply shows that in some cases, respect can be brought about via certain forms of manipulation. And this in turn suggests that in Plum’s case, even though the attitudes, deliberative tendencies, values, motives, intentions, and actions that constitute his disrespect of White are causally determined according to the neuroscientists’ plan, in treating White in the way that he does, Plum genuinely disrespects her standing as a fellow person.

***

These three arguments provide the compatibilist with powerful bases for claiming that in murdering White, Plum was disrespectful. And if I’m right to think that the blaming attitudes of resentment and indignation are made fitting when agents are disrespectful, then it follows that Plum is a fitting target of White’s loved ones’ indignation. Furthermore, because being a fitting target of blaming attitudes like resentment and indignation is what it is to be minimally morally responsible, this suggests that despite being the victim of manipulation, Plum is, in fact, a minimally morally responsible agent. Thus, hard compatibilism, defined as a thesis about minimal moral responsibility, appears to be vindicated.

---

Plausibly, if Kant is right, respect also gives Chum duties of beneficence. And so e.g., by failing to help her when she is in dire need, Chum would disrespect Knight. Of course, Chum is sufficiently respectful, so he not only allows her value as a person to serve as a side constraint on his actions, he also offers aid when such beneficence is required.
3.4.

But incompatibilists are not without a reply to this argument. In response, the incompatibilist might concede that Plum acts in a disrespectful way. But they might go on to argue that it is not mere disrespect that makes attitudes like resentment and indignation fitting. Instead, A’s blaming attitude towards B is fitting only if B freely disrespects A (or some other member of the moral community). And because Plum is manipulated to be disrespectful, he does not freely disrespect White. Therefore, the indignation that her loved ones would feel in response to his crime is not fitting.

In response, I would begin by distinguishing between two incompatibilist notions of “freedom” that might be at stake. First, the incompatibilist might be claiming that A’s blaming attitude towards B is fitting only if B is the causal-historical source of her disrespectful treatment of A. And second, she might be claiming that A’s blaming attitude towards B is fitting only if, at the time at which B disrespects A, B can refrain from disrespecting A. I consider each of these possibilities in turn.

First, the incompatibilist cannot simply assume that resentment is fitting (in the sense that it is deserved) only if the wrongdoer is the causal-historical source of her action. For why should we think this? Recall that the hard compatibilist about minimal moral responsibility and manipulation is takes no stand on whether manipulated agents can be apt targets of damnation, execution, and long prison sentences (at least when such sentences are justified in retributive terms). So it’s open for her to say, in response to hearing about Plum’s victimization at the hands of the neuroscientists, that he does not deserve to be tried and punished in a criminal court. Likewise the hard compatibilist can deny that Plum deserves to be condemned to eternal torture. In short, she can agree with Pereboom’s intuitions about such cases. However, it’s not clear that once she concedes this, Pereboom’s case retains its intuitive pull. For is it really intuitive that no one—not even White’s family and friends—can be indignant at his actions? Such an extreme view certainly seems implausible to me. To see this, consider the following scene from Margaret Atwood’s The Blind Assassin:

“Don’t be ridiculous,” said Winifred, though she looked shaken. “Richard’s hands are absolutely clean, whatever Laura said. He is pure as the driven. You’ve made a serious error
in judgment. He wants me to say he’s prepared to overlook this – this aberration of yours. If you’ll come back, he’s fully willing to forgive and forget.”

“But I’m not,” I said. “He may be pure as the driven, but it’s not the driven snow. It’s another substance entirely.”

“Keep your voice down,” she hissed. “People are looking.”

“They’ll look anyway,” I said, “with you dressed up like Lady Astor’s horse. You know, that colour of green doesn’t suit you one bit, especially at your present age. It never has, really. It makes you look bilious.”

This hit home. Winifred was finding it hard going: she wasn’t used to this new, viperish aspect of me. “What do you want, exactly?” she said. “Not that Richard did anything at all. But he doesn’t want an uproar.”

“I told him, exactly,” I said. “I spelled it out. And now I’d like the cheque.”

“He wants to see Aimee.”

“There is no way in Hell,” I said, “that I will permit such a thing. He has a yen for young girls…”

Here, the protagonist, Iris, has just confronted her sister-in-law (Winifred) about Iris’ estranged husband’s (Richard) repeatedly raping Iris’ sister (Laura), forcing her to have an abortion, and eventually, driving Laura to kill herself. Of course, as is the case in many, perhaps most, of our ordinary blaming interactions, Iris never says, “I blame thee…” Rather, she expresses her blaming attitudes—her contempt, resentment, and indignation in both what she says (e.g., literally comparing her husband’s purity to that of excrement, etc.) and how she says it (e.g., “I told him, exactly”).

However, suppose that Richard and Winifred had been manipulated in precisely the same way in which Plum is manipulated. Suppose that Richard was manipulated to rape Laura for perverse, self-absorbed reasons, and that Winifred was manipulated to aid him in hiding her subsequent pregnancy for egoistical reasons. Given the gravity of their crimes, of their blatant and despicable disregard for Laura and Iris—and really, anyone but themselves—is it really credible that, had they been so manipulated, Iris’s blaming attitudes would not have been fitting? I understand that, were they victimized in this way, that we should refrain from punishing Richard and

---

21 Margaret Atwood [2000] pgs. 505-06.
Winifred—though plausibly, we could isolate them in the ways that Pereboom advocates. But to conclude that they don’t deserve Iris’ blaming attitudes strikes me not only as wrong, but as inhumane. If we value Laura, her bodily integrity, and her right to decide for herself how best to live her life, blaming attitudes like those displayed by Iris would be entirely appropriate. And if this is right, then blaming attitudes like resentment and indignation can be fitting even when wrongdoers are not the causal-historical sources of their disrespect.

Similarly, if by “freely disrespecting” the incompatibilist means that at the time in which B disrespects A, B could have refrained from disrespecting A, she is in no better position. For suppose that though Richard and Winifred were not manipulated, they were monitored by a team of neuroscientists such that, were they to do the right thing and refrain from raping and abetting a rapist respectively, they would have been forced by the neuroscientists to act in those monstrous ways. We can further suppose that as it actually happens in the story, Richard and Winifred make their evil decisions “on their own” in such a way that the neuroscientists do not have to intervene. Of course, because of the neuroscientists’ presence, it seems that at the time of their evil decisions, neither Richard nor Winifred had the ability to refrain from acting in a way that was supremely disrespectful towards Laura. Nevertheless, for the very reasons I adduced above, it still seems to me that Iris was justified in her blame of Richard and Winifred. Indeed, again, it is not credible that Richard and Winifred do not deserve Iris’ contempt, resentment, and indignation. Thus, these blaming attitudes can be fitting even when wrongdoers cannot refrain from disrespecting their victims.

The hard compatibilist, then, has an answer to this incompatibilist challenge. To claim that blaming attitudes are not appropriate in cases like the one considered above seems morally insensitive. And if this is right, then in Plum’s case too, certain forms of blame are deserved. If we stopped White’s family and friends from expressing their blame, we would act wrongly—not, I think, because we lack the standing to interfere with their blame, but because their blame is, in fact, appropriate. Consequently, I conclude that it is disrespect itself that makes blaming attitudes fitting. And as a result, Plum is a fitting target of reactive attitudes like resentment and indignations and is therefore, minimally morally responsible for his crime.

22 See Pereboom [2001, forthcoming].
23 The structure of this case is due to Harry Frankfurt’s famous case. See Frankfurt [1969].
4. Manipulation and Interpersonal Relationships

Having argued that manipulation doesn’t undermine the fittingness of the reactive emotions, I now want to consider whether being causally determined by a team of neuroscientists to kill White precludes Plum from deserving to be targeted with an interpersonally significant form of distrust, a form of distrust that licenses significant alterations in how Plum may be treated by those with whom he interacts. The sorts of alterations I have in mind might involve, in cases like this, radical restructuring of existing relationships—a restructuring of relations that T. M. Scanlon has plausibly characterized as blame. However, if incompatibilism is correct, the presence of the neuroscientists in the causal history of Plum’s action renders it unfair to target him with these responses. But plausibly, I will argue, this form of blame is consistent with the presence of manipulation.

My argument for this claim will proceed in two stages: first I will argue that nothing about the manipulators’ presence in the causal history of Plum’s actions renders it unfair or inappropriate for us to distrust him. This argument parallels Pamela Hieronymi’s [2004] argument that lacking the ability to do otherwise does not undermine the force or fairness of moral blame. Second, because distrust is appropriate, it cannot be unfair for us to thereby restructure our relationships with Plum in a way that constitutes a relational form of blame. Thus, as we will see, Plum deserves to be blamed in the sense at stake.

4.1

For ease of exposition, let us begin by supposing that before the murder, Plum had been something of a friend to White, and that he was well-liked by her family and their mutual friends. Under this supposition, there is a standing relationship that obtains between Plum and his victim’s family. They are not simply strangers, only brought together in the wake of the murder of Ms. White. Instead, they have some antecedent degree of friendship. And the norms of governing their friendship require standing intentions to act beneficently towards each other, attitudes of good will towards one

24 There are good reasons to think that my arguments in this section will succeed even if we do not suppose that there is a standing friendship between White’s loved ones and Plum (beyond, of course, the moral relationship that we stand in with all persons). But for the sake of clarity, I work with this assumption in place. And in this respect, I’m simply following Scanlon’s development of his theory of blame [2008]. But note, in subsequent work, Scanlon has argued for a similar claim—that even if we don’t focus on the case of friendship, we can still interpret blame in distinctively relational ways. See Scanlon [2013].
another, and a set of normative expectations that circumscribe how they can demand to be treated as parties to such a relationship.

Now, consider how White’s family and friends might react when they learn that Plum has killed White for selfish reasons. Obviously, they would feel shocked and hurt. But in addition to the shock and pain that come from the death of a loved one, they would certainly feel betrayed. And probably, in light of his action, they would come to regard their relationship with Plum as being irrevocably broken. Not only is this response natural, it seems entirely appropriate given the gravity of Plum’s crime. Of course, by coming to regard their former relationship in this way, they thereby restructure their standing intentions, attitudes, and expectations towards Plum. For example, White’s family and friends might reasonably revise their standing intentions to show Plum good will. Furthermore, they would no longer regard Plum as someone whom they could trust. After all, in light of Plum’s willingness to disregard the standing of his friends, it is reasonable for them to withdraw their trust of him because he is someone with whom they can no longer engage in meaningful relations of shared respect and mutual recognition. That is, through his actions, Plum has shown himself to lack concern for this relationship and for the standards involved in maintaining such relationships in meaningful ways. And on the face of it, the withdrawal of trust seems warranted even though Plum was causally determined by his manipulators to murder White.

4.2

Pamela Hieronymi develops a closely related point when she claims that even in cases in which a person’s character is “due to formative circumstances outside of her control” (as is the case with Plum), it is not plausible “that those who interact with her can be charged with unfairness in distrusting her,” [Hieronymi 2004, 119]. Continuing this thought, Hieronymi claims that:

If a person will always let you down, that seems to be solid grounds for distrust rather than a condition under which distrust would be unfair. The distrust, it seems, simply marks the fact that a person’s unreliability or untrustworthiness is known. The force of the distrust is found in the importance of being, and being known to be unreliable or untrustworthy. If a person is unreliable or untrustworthy, that force seems rightly in place, [Hieronymi 2004, 199].
Hieronymi’s thought is that rather than it being unfair to distrust those, who, like Plum, come to be untrustworthy for reasons outside of their control, if their behavior legitimately makes it impossible, or even unreasonably difficult, for others to engage in meaningful relationships with them, then it can be appropriate for us to distrust them and for us to act in ways that reflect our distrust (e.g., by restructuring our relationships with these individuals).

Building on Hieronymi’s point, I want to claim that it isn’t mere unreliability as such that licenses distrust. The gravity of the breach of trust can similarly affect the appropriateness of distrust. After all, in many cases, it only takes one action to thoroughly undermine the possibility of continued relations, and such actions render distrust a fair response. For example, given the nature of many marriages, it often only takes one instance of infidelity to warrant distrust, even if the adulterous spouse has a long prior history of fidelity. Often, such breaches of trust are simply too destructive, and as a result, many relationships understandably do not survive it. This suggests that in some cases, it’s not mere unreliability that grounds the propriety of distrust.

And in the case of Plum, what licenses White’s family and friends’ withdrawal of trust not his history of unreliability. Rather, it’s the severity and irrevocability of his disrespect for Ms. White. Moreover, the restructuring of relations that attend the withdrawal of trust carry a characteristic force, viz., that this individual is someone with whom it’s not advisable, or even possible, to be involved with in meaningful reciprocal relations—relationships that carry with them an expectation of mutual trust. In other words, the force of distrust is that it licenses us to restructure our relationships with those who cannot be parties to relations of mutual trust.

But like Hieronymi, I want to emphasize that this force is not contingent on Plum being the ultimate causal source of his character or of his action. The presence of the manipulators doesn’t undermine Plum’s essential role as the agent who brings about White’s death. And this role is sufficient to license distrust of the sort that licenses a restructuring of relationships. So again, as in the case of blaming attitudes like resentment and indignation, even if White’s family and friends knew about the role of the neuroscientists in the causal history of Plum’s action, they would still be justified in restructuring their relationship with him on the basis of his action. Knowing that Plum was manipulated does not, by itself, render distrust unfair, and neither does it undermine the characteristic force of this distrust. Thus, if it is fair for White’s family and friends to distrust Plum,
and if a restructuring of their relationship with Plum is a fair consequence of their distrust, then *contra* the incompatibilist, it is not unfair for them to restructure their relationships with Plum.

4.3

But there is more we can say concerning the withdrawal of trust and subsequent restructuring of relations. It's plausible that when we respond to others' transgressions in ways that reflect the characteristic force of distrust, we are actually engaged in a morally significant form of blame. That is, a further thing we can say about the withdrawal of distrust and the subsequent restructuring of relationships is that these activities are plausibly thought of as a meaningful and significant form of blame—which is captured in T. M. Scanlon's interpretation of blame. More precisely, Scanlon claims:

Briefly put, my proposal is this: to claim that a person is *blameworthy* for an action is to claim that the actions shows something about the agent's attitudes toward others that impairs the relations that others can have with him or her. To *blame* a person is to judge him or her to be blameworthy and to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be appropriate [Scanlon 2008, 128-9].

And later, Scanlon supplements this proposal by adding that “to blame someone is to hold attitudes toward him that differ, in ways that reflect this impairment, from the attitudes required by the relationship one would otherwise have with the person,” [Scanlon 2008, 145].

Now obviously, I don't take Scanlon to have offered an exhaustive analysis of all the practices we have associated with blame and responsibility. There are other ways that we can respond to wrongdoing, ways that equally well deserve the title “blame” (e.g., his account of blame doesn't characterize the form of blame discussed in §3—a form of blame that is essentially tied to the reactive emotions like resentment and indignation). Nevertheless, I do think that Scanlon has provided us with an account of one important way we respond to others’ wrongdoing that is as worthy of the title “blame” as any other account on offer. Because of Plum's severe and irrevocable disrespect for Ms. White, her family and friends can legitimately withdraw their trust, and this withdrawal of trust has a characteristic force that leads them to radically restructure their relationship
with Plum. But notice, White’s loved ones are not simply giving up their relationship with Plum, they’re doing it on the basis of the interpersonal expressive significance of Plum’s action (i.e., the “meaning” of his action in Scanlon’s lingo)—an interpersonal expressive significance that, as I’ve already argued, obtains even though Plum has been causally determined to murder White by a team of neuroscientists. But the ending of this relationship is not relevantly similar to relationships that end or are impaired simply because one party moves away or is otherwise distant. As Angela Smith has recently put it, the marking of an impairment to a relationship and the restructuring that follows count as blaming insofar as these activities protest the meaning a wrongdoer’s action.\textsuperscript{25} Thus, in virtue of the fact that their distrust leads them to significantly restructure their relationship to Plum (a restructuring that counts as a kind of protest), White’s family and friends have blamed him for murdering White.

Of course, in light of their blame, the natural question to ask is whether such treatment is justified, i.e., whether Plum really is blame\textit{worthy} given the presence of the neuroscientists in the causal history of his action. But notice, as I’ve already argued (following Hieronymi), because this restructuring of relations with Plum is done on the basis of their warranted distrust, it is itself warranted. In other words, in light of his action, Plum deserves to be blamed in this way. Consequently, I think we should conclude that the mere fact that the team of neuroscientists has causally determined Plum to murder White doesn’t vitiate the hard compatibilist’s claim that Professor Plum is an apt target of the meaningful and interpersonally significant form of blame described by Scanlon.\textsuperscript{26}

\section*{5. Conclusion}
Taking stock, I have argued that despite the presence of the manipulators, Plum can plausibly be thought of as a fitting target of resentment and indignation. And if being minimally morally responsible is to be understood in terms of being a fitting target of the reactive emotions, then we should conclude that Plum is morally responsible for murdering White. I have also argued that it is

\textsuperscript{25} Smith [2013]. I should note that Smith takes herself to be adding to Scanlon’s interpretation of blame rather than merely expositing it. However, since this seems to be a plausible (and necessary) addition, I count it as part of the interpretation of blame under consideration.

\textsuperscript{26} Interestingly, Pereboom has recently admitted that blame, as it is interpreted by Scanlon, can sometimes be appropriate even if hard incompatibilism is true. For Pereboom, because this kind of blame does not entail that agents \textit{deserve} in the basic sense, to be treated harshly for their wrongdoing, it is not ruled out as potentially justified. See Pereboom [2013].
fair that we withdraw our trust from Plum, and that this withdrawal of trust involves a restructuring of our relationships with Plum—a relational restructuring that Scanlon has plausibly characterized as one form of blame. This suggests that even though Plum is the victim of manipulation, in virtue of the meaning of his action, it is fair for us to blame him. Thus, he is blameworthy. Perhaps, if this is right, then hard compatibilism isn’t so hard, and the manipulation argument is thereby an unsound reason to doubt the compatibility of minimal moral responsibility and causal determinism.
Works Cited


Mele, Al. [2006]. *Free Will and Luck*. Oxford University Press.


Russell and Oisin Deery. Oxford University Press.

Scanlon, T. M. [1998]. *What We Owe to Each Other*. Harvard University Press.


Todd, Patrick. [2012]. “Defending (a modified version of) the Zygote Argument,” *Philosophical Studies*.


