Defending Hard Incompatibilism Again[1]

Derk Pereboom, Cornell University

forthcoming in *Essays on Free Will and Moral Responsibility*, Nick Trakakis and Daniel Cohen, eds., Newcastle: Cambridge Scholars Press.

1. Hard incompatibilism characterized.

According to the hard incompatibilist position I advocate, we would not have the sort of free will required for morally responsibility if determinism were true. We would also lack this sort of free will if indeterminism were true and the causes of our actions were exclusively states or events. If the causes of our actions were exclusively states or events, indeterministic causal histories of actions would be as threatening to this kind of free will as deterministic histories are. However, it might well be that if we were undetermined agent-causes – if we as substances had the power to cause decisions without being causally determined to cause them – we would then have this sort of free will. But although our being undetermined agent causes has not been ruled out as a coherent possibility, it is not credible given our best physical theories. Thus we need to take seriously the prospect that we are not free in the sense required for moral responsibility (Pereboom 1995, 2001).

I oppose a type of incompatibilism according to which the availability of alternative possibilities is the most important factor for explaining moral responsibility, and accept instead a variety that ascribes the most significant explanatory role to the way in which the agent actually produces the action. In metaphysical terms, the sort of free will required for moral responsibility

[1] Thanks to Seth Shabo, Dana Nelkin, Michael McKenna, Ishtiyaque Haji, Louis deRosset, Randy Clarke, and David Christensen for very helpful commentary and discussion.

does not consist most fundamentally in the availability of alternative possibilities, but rather in the agent's being the causal source of her action in a specific way. Accordingly, I advocate *source* as opposed to *leeway* incompatibilism. Agent-causal libertarianism is typically conceived as an incompatibilism according to which an agent can be the causal source of her action in the way required for moral responsibility, and thus proponents of this view are typically source incompatibilists. But a source incompatibilist might seriously doubt that we have the sort of free will required for moral responsibility, and this is the position I defend.

But in addition, I contend that a conception of life without this type of free will would not be devastating to morality or to our sense of meaning in life, and in certain respects it may even be beneficial. The type of free will that is undermined according to the hard incompatibilism I advocate is the kind required for moral responsibility in the following specific sense: for an agent to be morally responsible for an action is for it to belong to her in such a way that she would deserve blame if the action were morally wrong, and she would deserve credit or perhaps praise if it were morally exemplary. The desert at issue here is basic in the sense that the agent, to be morally responsible, would deserve the blame or credit just because she has performed the action, given an understanding of its moral status, and not, for example, by virtue of consequentialist considerations, or solely by way of a contractualist account. This is the sense of moral responsibility that has been at issue in the debate about whether the sort of free will required for moral responsibility is compatible with determinism. Other notions of moral responsibility have not been at issue: for example, the legitimacy of calling agents to moral account, that is, the legitimacy of demanding that an agent explain how an action might be in accord with moral principles, and if this fails, of demanding that the agent take steps to avoid

similar behavior in the future.[2] The hard incompatibilism I advocate takes no issue with this notion of moral responsibility, or with the characteristics of agency required for it.

Philosophers not infrequently take on the task of rescuing ordinary beliefs and practices from threats that result from scientific or naturalistic conceptions of reality. Such conceptions have posed a challenge to belief in the sort of free will required for moral responsibility and to the attendant practice of holding people morally responsible; and also, for example, to belief in God, in an immaterial soul, in immortality, and to theistic religious practice. While naturalistic philosophers have often given up God, the soul, immortality, and religious practice, they have typically not come to deny moral responsibility in the sense at issue in the debate, or its attendant practice. In the phrasing of Wilfred Sellars (1963), they have not conceived of our moral responsibility, and the legitimacy of treating people as morally responsible, as a feature of the manifest image that has been undermined by the scientific image.

I argue that although denying that we are morally responsible in this sense has its cost to our ordinary self-conception, this cost is not as high as often thought. We would need to reject the rationality of basic desert, of the reactive attitudes that presuppose basic desert, of retributive justification of criminal punishment and personal recrimination, since all of this presupposes that we have the sort of free will required for moral responsibility in the sense at issue. What would survive untainted is the practice of calling each other to moral account, attitudes such as joy and sadness about what people do, justification for detaining criminals analogous to our rationale for quarantining carriers of dangerous diseases, and enjoyment of our achievements on a par with our enjoyment of our natural gifts. If we are careful to separate what in our conception of

[2] Arthur Kuflik suggested this notion of responsibility to me in conversation

morality and meaning in life is undercut by naturalism and what is not, we will see that we can live with what remains (Pereboom 1995, 2001).

2. A defense of the "Tax Evasion" Frankfurt-style case.

Why opt for a source as opposed to a leeway position? I argue that examples of the kind devised by Frankfurt yield an effective challenge to the leeway position (Frankfurt 1969). In those examples, an agent considers performing some action, but an intervener is concerned that she will not come through. So if she were to show some sign that she will not or might not perform the action, the intervener would arrange matters so that she would perform it anyway. Here is one of John Fischer's examples: Jones will decide to kill Smith only if Jones blushes beforehand. Jones's failure to blush (by a certain time) can then function as the prior sign that would trigger the intervention that would cause her to kill Smith. Suppose that Jones acts without intervention. Here we might well have the intuition that she is morally responsible for killing Smith, even though she could not have done otherwise than to kill Smith, and even though she could not even have formed an alternative intention. She could have failed to blush, but Fischer argues that such a flicker of freedom is of no use to the libertarian, since it is not sufficiently *robust* to have a role in grounding the agent's moral responsibility (Fischer 1994, 131-59).

Here is my earlier (2000; 2001, 26) proposal what it is for an alternative possibility to be robust:

> Robustness (1): For an alternative possibility to be relevant per se to explaining an agent's moral responsibility for an action it must satisfy the following characterization:

she could have willed something other than what she actually willed such that she understood that by willing it she would thereby have been precluded from the moral responsibility she actually has for the action

The intuition that underlies the proposal to ground moral responsibility in the accessibility of alternative possibilities is of the following sort: to be blameworthy for an action, the agent must have been able to do something that would have precluded her from being blameworthy, at least to the degree she's blameworthy (Pereboom 2001, 1). Accordingly, for an alternative possibility to be robust, it must first of all satisfy this condition: she could have willed something other than what she actually willed such that by willing it she would thereby have been precluded from the moral responsibility she actually has for the action (cf. Otsuka 1998). Secondly, the epistemic element of Robustness (1) – that she must have *understood* that by willing otherwise she would have been precluded from the responsibility she actually has – is motivated by the following sort of consideration. Suppose that that the only way Joe could have avoided deciding to take an illegal deduction on his tax form -- a choice he does in fact make -- is by voluntarily taken a sip from his coffee cup, for unbeknownst to him, the coffee was laced with the drug that induces compliance with the tax code. In this situation, he could have behaved voluntarily in such a manner that would have precluded the choice for which he was in fact blameworthy, as a result of which he would have been morally non-responsible for it. But whether he could have voluntarily taken the sip from the coffee cup, having no understanding that it would render him blameless in this way, is intuitively irrelevant to explaining whether he is morally responsible for his choice.

But here are two concerns for Robustness (1):

(a) One might imagine an agent who has alternative possibility, where so acting would preclude the responsibility she has for the option she selects, but due to some epistemic failing on her part, she does not believe that she has an alternative possibility that meets this specification. Dana Nelkin (in correspondence) suggests a case in which an agent mistakenly believes that the alternative possibility does not preclude the responsibility she has for the option she selects, but she does recognize significant morally salient differences between the two options. One might propose that the agent has a robust alternative possibility partly because there are good reasons available to her for believing that she has an alternative in which her responsibility is different in the relevant way, even though she does not appreciate those reasons adequately, but only partially. But, first, imagine that Joe should have known what effect drinking the coffee would have, because he should have been paying attention when this fact about the coffee was revealed at his Tax Evaders Anonymous class. Does he, as a result, now have a robust alternative possibility? Not clearly, and I would say not. Note that denying that he has a robust alternative possibility does not preclude the advocate of a principle of alternative possibilities from assessing him as derivatively responsible for evading taxes, for the reason that he may have met a relevant epistemic condition on derivative moral responsibility when he neglected to pay attention in the class. In addition, I'm inclined to deny that such an epistemic failing supplemented by a mere partial understanding of morally salient differences between accessible alternatives is enough for robustness, given that the partial understanding does not amount to an understanding that availing herself of an alternative possibility would preclude the responsibility she actually turns out to have. Suppose that Suzy could have saved Billy from a painful death by giving him an additional injection, but that she has no understanding of this since she wasn't

paying enough attention to the instructions when she should have been. But she was paying enough attention to understand that Billy would have been more comfortable had she given him the injection. My sense is that she does not have a robust alternative possibility in this case that would ground moral responsibility for allowing Billy to die, but still that she is perhaps responsible for allowing Billy to die derivatively from her not paying attention when she should have been – depending on the details of the case.

(b) In this example, is having a non-occurrent or even occurrent belief that taking a sip from the coffee cup *might* result in his not evading taxes enough for robustness (Ginet 2000)?[3] It seems not. For, if asked, Joe might well agree that the probability of this connection is non-zero – he might admit, for instance, that it's at least .000001, and if he's taken a class is epistemology or probability, something like this might well be his response. But, intuitively, this is not sufficient to generate robustness. Should it be required for robustness that Joe understood that taking the sip of coffee would, with a probability of 1.0, result in his not evading taxes? This is clearly too strong, for it would intuitively be enough for robustness if he understood that the probability was, say, .95.[4] But the threshold probability, as one would expect, is difficult or impossible to determine. So here is my new proposal:

> Robustness (2): For an alternative possibility to be relevant to explaining why an agent is morally responsible for an action it must satisfy the following characterization: she could have willed something different from what she actually willed such that she understood

[3] Kevin Timpe defended such a condition in the presentation of his paper "How Troublesome is Tracing" at the *Responsibility, Agency, and Persons* conference at the University of San Francisco in October, 2007.

[4] Jonathan Vance made this point in conversation, and Kevin Timpe argued in his presentation at the conference in San Francisco in October 2007 (see the previous note).

that by willing it she would be, or at least would likely to be, precluded from the responsibility she actually has.

Perhaps the most significant objection that has been raised against the earlier kinds of Frankfurt-style arguments was initially suggested by Robert Kane and then systematically developed by David Widerker and Carl Ginet (Kane 1985, 51; 1996, 142-4, 191-2; Widerker 1995, 247-61; Ginet 1996). The general form of the Kane/Widerker/Ginet objection is this: for any Frankfurt-style example, if universal causal determinism is assumed to hold in the actual causal sequence that results in the action, the libertarian will not have and cannot be expected to have the intuition that the agent is morally responsible. If, on the other hand, libertarian indeterminism in this actual sequence is presupposed, the scenario will not serve the Frankfurt-defender's purpose, for any such case will fall to a dilemma. In Frankfurt-style cases the actual situation will feature a prior sign that signals the fact that intervention is not required. If in the proposed case the prior sign causally determined the action, or if it were associated with some factor that did, the intervener's predictive ability could be explained. However, then the libertarian would not and could not be expected to have the intuition that the agent is morally responsible. But if the relationship between the prior sign and the action were not causally deterministic in such ways, then it will be the case that the agent could have done otherwise despite the occurrence of the prior sign. Either way, an alternative-possibilities condition on moral responsibility emerges unscathed.

I have proposed a Frankfurt-style scenario that avoids this objection (Pereboom 2000; 2001, pp. 18-19; 2003). Its distinguishing features are these: the cue for intervention must be a

necessary rather than a sufficient condition, not for the action that the agent actually performs, but for the agent's availing herself of any robust alternative possibility (without the intervener's device in place), while the cue for intervention itself is not a robust alternative possibility, and the absence of the cue for intervention in no sense causally determines the action the agent actually performs. Here is the example:

> *Tax Evasion (2*): Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He knows that claiming this deduction is illegal, but that he probably won't be caught, and that if he were, he could convincingly plead ignorance. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. In addition, the only way that in this situation he could fail to choose to evade taxes is for moral reasons, of which he is aware. He could not, for example, choose to evade taxes for no reason or simply on a whim. Moreover, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to moral reasons. Joe can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, he could, exercising his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). However, to ensure that he will choose to evade taxes, a neuroscientist has, unbeknownst to Joe, implanted a device in his brain, which, were it to sense the requisite level of attentiveness, would electronically stimulate the right neural centers so as to inevitably result in his making this choice. As it happens, Joe does not attain this

level of attentiveness to his moral reasons, and he chooses to evade taxes on his own, while the device remains idle. (David Hunt also suggests this 'necessary condition" strategy (2000), and develops a similar example (2005)).

In this situation, Joe could be morally responsible for choosing to evade taxes despite the fact that he could not have chosen otherwise.

The example does feature alternative possibilities that are available to the agent -- his achieving higher levels of attentiveness to moral reasons. Indeed, at this point one might object that given that the intervener's device is in place, by voluntarily achieving the specified higher level of attentiveness Joe would have voluntarily done something whereby he would have avoided the blameworthiness he actually incurs (Otsuka 1998). For had he voluntarily achieved the requisite level of attentiveness, the intervention would have taken place, whereupon he would not have been blameworthy for deciding to evade taxes. But this alternative possibility is not robust. Joe does not understand, and, moreover, he has no reason to believe, that voluntarily achieving the requisite level of attentiveness would or would likely preclude him from responsibility for choosing to evade taxes. True, were he voluntarily to achieve this attentiveness, the intervention would take place, and he would then not have been responsible for this choice. Still, Joe has no inkling, and has no reason to believe, that the intervention would then take place, as a result of which he would be precluded from responsibility for this choice. In fact, one might imagine that he believes that achieving this level of attentiveness is compatible with his freely deciding to evade taxes anyway, and that he has no reason to suspect otherwise. Nevertheless, Joe is morally responsible for deciding to evade taxes.

3. Robert Kane's challenge to Tax Evasion.

Kane's reply to Tax Evasion (Fischer, Kane, Pereboom, and Vargas 2007, 171) crucially features the claim that the controller "is not going to let Joe make the undetermined choice between A and B," where A is the choice to evade taxes, and B is doing otherwise, and from this Kane concludes that Joe will not be (non-derivatively) morally responsible for the choice to evade taxes. His argument is this: if the cue for intervention, Joe's attaining the requisite level of attentiveness to moral reasons, does not occur, and he thus chooses A since the necessary condition for choosing B is not in place, Joe's decision "will not be a "will-setting" SFA (self-forming action)... because he will only have reasons to "set his will" on A and will not have attended to any good reasons to set his will on B." If he does attain the level of attentiveness, the controller will intervene and make him choose A, and "so Joe will not get a chance to make a true SFA *either way* once the controller is in the picture."

Thus the reason Kane cites for Joe's not being non-derivatively morally responsible is that he will not have the undetermined choice between A and B. Notice that he is contending that Joe is not morally responsible because he cannot do otherwise. More precisely, Kane is claiming that Joe is not responsible because he lacks plural voluntary control, and in the sense specified by this notion, a robust alternative possibility. However, this is just the issue the leeway and the source theorist are arguing about, i.e., whether robust alternative possibilities are required for moral responsibility. In order to advance the debate, the source theorist devises a Frankfurt-style case in which the agent lacks robust alternative possibilities, but which is intended to elicit the intuition that he is morally responsible. What are we then to say of the response that the agent is not responsible because he lacks robust alternative possibilities?

It would be mistaken to say that Kane's response actually begs the question against the Frankfurt-defender.[5] For the success of a Frankfurt-style argument depends on whether the audience finds it intuitive that the agent is morally responsible. As it turns out, Kane does not find it intuitive that Joe is morally responsible. For him, the ultimate reason is that Joe lacks alternative possibilities, and this view may, in the last analysis, be correct. Still, there is a respect in which this response to a Frankfurt-style case is unsatisfying, since it explicitly cites the leeway position on what is at issue as the reason why Joe is not morally responsible. To be sure, one can run a principle of alternative possibilities through any example, and then tally the results. But this procedure stands to miss the force of what might be a counterexample, and thus runs a serious risk of failing to engage an objection. Accordingly, it is at least prima facie dialectically unsatisfying. Moreover, this procedure precludes he possibility of discussing the issue at hand by way of Frankfurt-style cases. For we know in advance what, ultimately, the response to any such case will be: the agent is not responsible because he lacks robust alternative possibilities.

Like many philosophical discussions, the correct outcome of the debate about the principle of alternative possibilities should be viewed as a matter of reflective equilibrium.[6] On Frankfurt's side, we have the intuitions about examples, such as Tax Evasion, that skirt the Kane/Widerker/Ginet objection. On the other side, there is the force of Widerker's W-defense: About an agent who breaks a promise, but could not have done otherwise he writes:

---

[5] I respond to Kane in (Fischer, Kane, Pereboom, and Vargas 2007, 191-4), but what follows is a more considered evaluation. Among other things, I no longer believe that Kane is merely superimposing a PAP-schema on my case. Rather, as I argue even there, his analysis is backed by the regress of motives argument for an alternative-possibilities requirement.

[6] On the advisability of thinking of such dialectical situations as a matter of reflective equilibrium, see Sommers (forthcoming).

> Still, since you, [Harry] Frankfurt, wish to hold him blameworthy for his decision to break his promise, tell me *what, in your opinion, should he have done instead*? Now, you cannot claim that he should not have decided to break the promise, since this was something that was not in his power to do. Hence, I do not see how you can hold Jones blameworthy for his decision to break the promise. (Widerker 2000, 191)

The Frankfurt-defender can point to no alternative possibility, and must instead focus attention on how the agent actually did behave. Here I think that Michael McKenna has it right (2005, 177). When Widerker asks of Joe, in view of the fact that he had no robust alternative possibility: what would you have him do? one should admit that there is no good answer. But instead we should call attention to what Joe has actually done, and to the causal history by which his action came about. Moreover, this case should not be assimilated to one in which Joe acts in some sense against his will because he has only one genuine option for action. He is a wholehearted tax evader; we might even set up the case so that he wills to evade taxes, he wants to will to do so, and he wills to evade taxes because he wants to will to do so.

Kane's concern is somewhat different. His argument for requiring plural voluntary control is to be found in the "motives" part of his dual regress argument (Kane 1996, 2000). There he contends that for an agent to set her will requires that she have access to what are in effect robust alternative possibilities. As Seth Shabo (forthcoming) points out, Kane's concern here is whether the motivations present in a situation provide decisive reasons to choose as he does, and in a controversial case, one needs to ask whether the agent's will is set in this way. Kane contends that if the agent's will is set by the motivations present in the situation, then non-derivative moral responsibility is precluded. Now one might argue that if there is no actual

conflict of motivations for the agent, as is the case in Tax Evasion, then his will is set in the non-derivative-responsibility-precluding way at issue. But this does not seem right. For even though there is no actual conflict of motivations for Joe, solely by way an exercise of his libertarian free will he could have been more attentive to the moral reasons, whereupon the conflict could have ensued. True, he would have had to be more attentive to these moral reasons than he actually was in order for them to motivate him, but nothing about the situation prevents him from achieving this level of attentiveness. So, intuitively, his will was not set in the non-derivative-responsibility-precluding way by motivations present in the situation, or by anything else about the situation. In Shabo's phrasing, although Joe's will is perhaps provisionally set, it is not conclusively set, since he can voluntarily achieve the attentiveness that makes not evading taxes a genuine possibility (Shabo forthcoming). So what remains on the other side is the W-defense, which has significant force, but so does the intuition of moral responsibility in the successful Frankfurt-style cases. In addition, the Frankfurt-defender has a response to the W-defense, while if Tax Evasion works, it seems that the advocate of the principle of alternative possibilities has no response of equal or greater strength.

4. David Widerker's response to Tax Evasion.

Widerker has recently developed a challenge to Tax Evasion that can be viewed as a filled-out version of Kane's (Widerker 2006).[7] The idea is to apply the distinction between non-derivative and derivative responsibility to the example, where non-derivative moral responsibility

[7] What follows is a reply to one of Widerker's two criticisms of Tax Evasion; the second is a timing worry that echoes some of Carl Ginet's concerns (1996, 2002), to which I respond in (2001, 28-33) and in (2005).

is subject to a principle of alternative possibilities (PAP). Widerker, unlike Kane, argues that Joe is derivatively blameworthy for his decision to evade taxes:

> Another problem with Pereboom's example is that, in it, the agent is *derivatively* blameworthy for the decision he made, because he has not done his reasonable best (or has not made a reasonable effort) to avoid making it. He should have been more attentive to the moral reasons than he in fact was – something he could have done. And in that case, he would not be blameworthy for deciding to evade taxes, as then he would be forced by the neuroscientist so to decide. If this is correct, then Pereboom's example is a case of derivative culpability, and hence is irrelevant to PAP, which… concerns itself only with direct or nonderivative culpability. (Widerker 2006, 173)

Again, there is a sense in which this is a dialectically unsatisfying response to a Frankfurt-style case, since it explicitly cites a leeway position in justifying its claims about Joe's responsibility. Joe is non-derivatively responsible only for not deciding to be more attentive to the moral reasons, for only at this point is an alternative possibility available to him, and any responsibility he has for deciding to evade taxes must be derivative of this earlier decision. To be sure, the following schema can indeed be applied to any example in which moral responsibility is at issue: the agent is non-derivatively responsible at some particular time only if alternative possibilities are accessible to her at that time, or if robust alternative possibilities are accessible to her at that time, and all other responsibility is derivative from such non-derivative responsibility. But in the case of Tax Evasion, the concern is that the force of the example is not being engaged. One can of course run this PAP-schema through any example, and then note the result. But this procedure stands to miss the force of a potential counterexample, and hence risks failing to

engage a serious objection.

Still, this PAP-schema may be correct. However, in discerning whether it might be, we need to examine the drawbacks for imposing it on situations like Joe's. One might initially think that there will be little or no concern, for one's intuitions about whether agents are morally responsible do not distinguish between non-derivative and derivative responsibility. So it may be intuitive that Joe is morally responsible for deciding to evade taxes, while it is not intuitive that he is non-derivatively as opposed to derivatively responsible. However, the paradigm example of derivative responsibility is of the following sort: an agent decides to get drunk, understanding that when he is intoxicated he will not be able to avoid being abusive to his companions. In this case, while the general conditions on moral responsibility – that is, on non-derivative moral responsibility – uncontroversially fail to hold when he is drunk, nonetheless they do hold at the point when he decides to get drunk. However, this agent's situation differs significantly from Joe's. Our intoxicated agent has knowingly put himself in a position in which the general conditions on non-derivative moral responsibility uncontroversially fail to hold at subsequent times. This is not true of Joe, since when, at any given time, he fails to be sufficiently attentive to moral reasons, he understands that it remains open to him to become more attentive at subsequent times. Consequently, Joe's situation differs in a crucial respect from the paradigm case of derivative responsibility. He never knowingly puts himself in a position in which the general conditions on non-derivative moral responsibility uncontroversially fail to hold. To my mind, this strongly indicates that the application of Widerker's PAP-schema to Joe's case in the way suggested by the objection is ruled out.

Finally, in his critical analysis of Tax Evasion, Widerker argues: "he should have been

more attentive to the moral reasons than he in fact was – something he could have done. And in that case, he would not be blameworthy for deciding to evade taxes, as then he would be forced by the neuroscientist so to decide" (Widerker 2006, 173). All of this is true, but it is not enough to make the alternative possibility that is available to him robust relative to responsibility for deciding to evade taxes, since Joe has no sense at all that becoming more attentive to the moral reasons would result in his being forced to make this decision, and hence not blameworthy for doing so. Moreover, given the set-up of the case, it is flase that Joe should have had even the slightest inkling that becoming more attentive would have this result.

5. John Fischer's argument that the earlier sorts of Frankfurt-style cases are effective.

In response to the Kane/Widerker/Ginet objection, Fischer has advanced a subtle claim about the dialectical structure of the discussion of Frankfurt-style arguments, whose upshot would be that even early Frankfurt-style cases, like his blush example, would have significant force against a leeway position. Then Frankfurt-style cases that were not constructed with Kane/Widerker/Ginet objection in mind, and thus did not take care to avoid causal determinism in the actual sequence, would be effective, and the need for examples, like Tax Evasion, which were designed to answer this objection, would not be pressing. Fischer contends that earlier cases, even if they assume causal determinism in the actual sequence, nonetheless indicate that if the agent is not morally responsible, this is not simply because she could not have done otherwise:

> I think that the examples make highly plausible the preliminary conclusion that if Jones is not morally responsible for his choice and action, this is not simply because he lacks

> alternative possibilities. After all, everything that has causal (or any other kind of) influence on Jones would be exactly the same, if we "subtracted" Black [the intervener] entirely from the scene. And Jones's moral responsibility would seem to be supervenient on what has an influence or impact on him in some way. So the relevant (preliminary) conclusion is, if Jones is not morally responsible for his choice and action, this is not simply because he lacks alternative possibilities. And it does not appear to beg the question to come to this conclusion, even if causal determinism obtains. (1999, 113)

I agree that such early Frankfurt-style arguments definitely enliven the hypothesis that facts about an action's actual causal history, rather than alternative possibilities, are the key factor in explaining an agent's moral responsibility, and that as a result these early examples provide some reason to believe that these facts indeed have this explanatory role. But how decisive are these early arguments? The answer might be relative to one's initial position in the debate, as Ishtiyaque Haji and McKenna have also contended (2004; see also Pereboom 2003, 190-3). It might be that a Frankfurt-style argument of this early sort is more decisive for someone who is initially a leeway compatibilist than for someone who is at first a leeway incompatibilist. Suppose it turns out that in an early Frankfurt-style example, the actual causal history that produces the action is in fact causally deterministic, but this determinism is not responsibility-undermining given compatibilist intuitions. This example might then provide a leeway compatibilist with a strong reason to abandon her view, in favor, say, of Fischer's kind of compatibilism (1994, 1998), which features an actual causal history account of moral responsibility. However, it is not nearly so clear that this example would provide as strong a reason either to a leeway incompatibilist or an uncommitted participant in the debate who is

nevertheless concerned that determinism all by itself precludes moral responsibility.

The key issue here is whether, in the Frankfurt-style example at issue, in the last analysis it is causal determinism in the actual sequence that explains the absence of alternative possibilities, and in particular, whether, perhaps despite initial appearances, the preclusion of alternative possibilities by way of the counterfactual intervention is dependent on causal determinism in the actual sequence. If the example is so deeply dependent on causal determinism in the actual sequence, then it would provide not even the uncommitted observer who is nonetheless concerned that determinism might rule out alternative possibilities with a strong reason to abandon a principle of alternative possibilities. Stewart Goetz has pressed this point, arguing that "with determinism in the actual sequence it is not the device that prevents Jones from making an alternative choice," and the appearance that the intervener's device has this role is an illusion (Goetz 2005, 85; 2002).

In his response to Goetz's claim, Fischer agrees that in these examples causal determinism in the actual sequence *is one of the factors* that explains the agent's lack of alternative possibilities.[8] However, he argues that these examples feature two factors that explain the absence of alternative possibilities: the first is the causally deterministic process of the actual sequence that produces the action, and the second is the non-actual process involving the intervener. He then points out that we can consider these two action-ensuring conditions separately. Indeed, we can bracket the causally deterministic process of the actual sequence, and focus just on the non-actual process involving the intervener. This non-actual process ensures the action in question, and yet we have the intuition that an agent can be morally responsible

[8] What follows is a streamlined and perhaps improved version of the criticism I develop in (Pereboom 2006, section 3)

despite the conditions of this process being in place. Or, more cautiously, we at least have the intuition that if the agent is not morally responsible, it is not because the intervener's presence rules out alternative possibilities. An agent might then be morally responsible despite action-ensuring conditions being in place that do not depend on determinism in the actual sequence to preclude alternative possibilities. Thus the uncommitted participant has a strong reason deriving from this example to believe that if the agent is not morally responsible, it is not because the intervener's presence rules out alternative possibilities:

> In the Frankfurt-type scenarios, *two* causes make it the case that Jones is unable to choose otherwise at T2: the prior condition of the world (together with the laws of nature) and Black's counterfactual intervention. What these examples show is that the mere fact that Jones is unable to choose otherwise does not in itself establish that Jones is not morally responsible for his choice. This is because Black's counterfactual intervention is one of the factors that make it the case that Jones is unable to choose otherwise at T2, and yet it is irrelevant to the grounding of Jones' moral responsibility. Considering this factor (the counterfactual intervention), and bracketing any other factor that might make it the case that Jones is unable to choose otherwise at T2, it seems to me that Jones may well be morally responsible for his [choice]. The mere fact that he lacks alternative possibilities, then, cannot in itself be the reason that Jones is not morally responsible, if indeed he is not morally responsible. (Fischer 2006, 199-200)

In developing his claim that it is an illusion that the intervener's device precludes alternative possibilities, Goetz contends that in general, the absence of alternative possibilities can only be explained by causal determinism in the actual sequence (2005, 91-2). However, there are two

ways of thinking about determinism: (a) as involving the claim that events are *entailed* by propositions that describe preceding conditions and the laws of nature, and (b) as involving the claim that events are actually *produced* by such preceding conditions the laws of nature. A proposition describing the conditions in the set-up of a suitably constructed Frankfurt-style case entails that the action will come about, and so the action will then be determined by these conditions in sense (a). But since all of these conditions are not actually operative -- some are merely counterfactual -- it still seems open that just by virtue of a counterfactual intervener ensuring the action, such a case does not feature causal determinism in sense (b). Terminologically, to my ear sense (b) is genuine *causal* determination by preceding conditions, while sense (a) involves determination by preceding conditions that need not amount to causal determination. In Tax Evasion, I think I've produced a Frankfurt-style case in which an action is ensured so that the preceding conditions determine the action in sense (a) without causally determining them (sense b).[9] It has not been ruled out that determination in sense (a) is impossible without causal determination, but I think that the case has not yet been made.

I contend that we should at least provisionally grant the distinction between (a) and (b), and that Fischer's analysis is best challenged by a different tack. Fischer's account relies on the existence of two separate sets of ensuring conditions, one of which involves the merely counterfactual intervention, while the other consists just in the causally deterministic process in the actual sequence that produces the action. Let us allow, as Fischer does, that the causally

[9] Joseph Campbell (2006) in effect points out that source incompatibilists who claim that non-responsibility transfers through determination in sense (a) cannot at the same time appeal to standard Frankfurt-style cases to rule out the leeway position, since in those cases features of the set-up determine the action in sense (a). According to my source incompatibilism, non-responsibility transfers through determination in sense (b), but not in sense (a)

deterministic process in the actual sequence that produces the action genuinely amounts to a first set of ensuring conditions. But now focus on the second set of ensuring conditions, which features the counterfactual intervener. The key to the example's being a successful Frankfurt-style case is that this second set must in fact be a set of ensuring conditions. However, I think that in the earlier type of Frankfurt-style cases (not in Tax Evasion, for instance), this second set would not be a set of conditions that ensures the action *unless the actual causal sequence was deterministic*. For in such examples, the presence of the counterfactual intervener, all by itself, that is, independently of what happens conditionally on the prior sign occurring, does not ensure that the action will take place. In Fischer's "blush" example, the intervener's ensuring the action is conditional on the blush not occurring by a specified time. But in the set-up of the case, it is also possible that the blush does occur by the specified time. What produces the action if the blush does occur is what happens in the actual causal sequence. And as Widerker has pointed out, what happens in the actual sequence must be causally deterministic relative to the blush's occurring if the action is to be guaranteed. If it weren't causally deterministic relative to the blush's occurring, then the actual causal sequence would feature alternative possibilities subsequent to the blush. Thus, the second ensuring condition, the one involving the counterfactual intervener, would not be an ensuring condition unless the actual causal sequence were deterministic.

We saw that Goetz contends that with determinism in the actual sequence it is not the device that prevents Jones from making an alternative choice, and the appearance that the intervener's device has this role is an illusion. We can now see that this might well be correct for these earlier Frankfurt-style cases. The most dramatic and attention-diverting feature of such a

story, the presentation of the facts about the intervener, leads the audience to believe that these facts all by themselves function as an ensuring condition. But more careful analysis reveals that these facts would not amount to an ensuring condition if the actual sequence that produces the action were not causally deterministic.

In summary, the force of a Frankfurt-style case lies in the fact that the action is ensured while it is nevertheless intuitive that the agent is morally responsible. To achieve this guarantee in the path involving the counterfactual intervener, the blush example requires causal determinism in the sequence leading from the blush to the action. If the blush occurs, the intervention will not, so then any guarantee of the action that the presence of the intervener might provide is no longer in play. In this event, the blush's occurring must guarantee the action, otherwise the example will feature alternative possibilities after all. Thus while this sort of case may involve two distinct sets of ensuring conditions, the problem is that each set must feature, crucially, an actual sequence that is causally deterministic. This fact will undermine the force of the example for salient audiences. A Frankfurt-style case of this kind will not be especially effective in providing the uncommitted participant or the incompatibilist strong reason to reject a leeway position. But this is not a fatal problem for Frankfurt-style cases generally, since there are the more recently developed examples, like Tax Evasion (2000, 2001, 2003), a similar case developed by Hunt (2005), perhaps the Al Mele-David Robb scenario (1999; Mele 2006), and McKenna's (2003) example that avoid causal determinism in the actual sequence.[10]

7. A defense of the four-case manipulation argument against compatibilism.

[10] Haji and McKenna (2006) argue that Tax Evasion has dialectical force if modified as a deterministic example.

If Frankfurt-style cases are successful, both compatibilist and incompatibilist versions of the source position remain as live options. According to source compatibilism, an agent's moral responsibility for an action is to be explained not by the availability to her of alternative possibilities, but by the action's having a causal history of a sort that allows her to be the source of her action in a specific way, and compatibilism is true. Fischer is an advocate of a view of this kind, and he is thus an opponent of source incompatibilism. While he noted the possibility of source incompatibilism early on (Fischer 1982), he has argued that "there is simply no good reason to suppose that causal determinism in itself (and apart from considerations pertaining to alternative possibilities) vitiates our moral responsibility" (Fischer 1994, 159; 2006, 131, 201-2).

I think that the best type of challenge to the compatibilist at this point develops the claim that an action's being produced by a deterministic process that traces back to factors beyond the agent's control, even when she satisfies all the conditions on moral responsibility specified by the contending compatibilist theories, presents in principle no less of a threat to moral responsibility than does deterministic manipulation. My "four-case argument" first of all features examples that involve such manipulation, in which the agent satisfies these compatibilist conditions on moral responsibility, and which elicit the intuition that she is not morally responsible. In particular, the argument sets out three such cases, each progressively more like a fourth scenario, one that the compatibilist would count as realistic, in which the agent is causally determined to act in a way that is uncontroversially natural. The challenge to the compatibilist is to point out a difference between the manipulation examples and the fourth, ordinary, scenario that shows why the agent can be morally responsible in the ordinary case, and not in one or more of the manipulation examples. My contention is that non-responsibility generalizes from at least one of

the manipulation cases to the fourth, ordinary one.

In each of the four cases, Professor Plum decides to kill and does in fact kill Ms. White for the sake of a personal advantage. His action conforms to the prominent compatibilist conditions, which are designed to be sufficient for an agent's moral responsibility when supplemented by some fairly uncontroversial additional necessary conditions. First, it satisfies the various conditions proposed by Hume and his followers: the action is caused by desires that flow from his "durable and constant" character, since for him egoistic reasons typically weigh very heavily -- much too heavily as judged from the moral point of view, while the desire on which he acts is nevertheless not in some ordinary sense irresistible for him, and in this respect he is not constrained to act (Hume 1739/1978, 319-412). Next, it fits the condition proposed by Frankfurt: Plum's desire to murder White conforms to his second-order desires in the sense that he wills to murder her and wants to will to do so, and he wills to perform this action because he wants to will to do so (Frankfurt 1971). The act satisfies the reasons-responsiveness condition proposed by Fischer and Ravizza: for instance, Plum's desires are modified by, and some of them arise from, his rational consideration of the reasons, he is receptive to the appropriate pattern of reasons, and if he understood that the bad consequences for himself that would result from killing Ms. White would be significantly more severe than he judges them likely to be, he would have refrained from killing her for this reason (Fischer and Ravizza 1998, 69-82). His action also meets a condition proposed by Jay Wallace: while he deliberates and acts, Plum has the general capacity to grasp, apply, and regulate his behavior by moral reasons. For example, when egoistic reasons that count against acting morally are relatively weak, he will typically regulate his behavior by moral reasons instead. These capacities even provide him with the ability to revise

and develop his moral character over time (Wallace 1994, 51-83). Now, given that causal determinism is true, is it plausible that Plum is responsible for his action?

In a first type of counterexample (Case 1) to these prominent compatibilist conditions, the manipulation is local – sophisticated neuroscientists manipulate Plum from moment to moment by radiotechnology so that he is thereby causally determined to act in such a way that the compatibilist conditions are met. One might specify that the manipulation takes place at every moment, and directly affects Plum at the neural level, with the result that his mental states and actions feature the psychological regularities and counterfactual dependencies that those of an ordinary agent might exhibit. Hereby the incompatibilist aims to elicit the intuition that Plum is not morally responsible, thus showing that the prominent compatibilist conditions are not sufficient for moral responsibility.

This example might be filled out in response to those, such as Fischer, Mele, and Lynne Baker, who have wondered whether Plum in Case 1 meet certain minimal conditions of agency because, for example, that he is too disconnected from reality, or that he himself lacks ordinary agential control (Fischer 2004, 156; Mele 2005, 78; Baker 2006, 320). McKenna suggests a way of countering this kind of move: "Let us instead assume that Team Plum [the team of manipulators] operates by providing a very weird causal prosthetic, a causal foundation of *Plum's* control (i.e., a foundation different from the foundation provided by typical neural realizers found in normal agents)" (McKenna, forthcoming). We might suppose that many of Plum's mental states are realized in the usual neural way, but some are partially realized by the activities of the manipulators. Or else -- and this is how I prefer to set up the case -- the manipulators can, and sometimes do, locally manipulate Plum's neural states, his mental states

are realized in the ordinary way by his neural states, and since the psychological results of this manipulation cohere with Plum's non-manipulated motivation and action, the upshot is an agent with an internally coherent character. In addition, I specify in the set-up of Case 1 that Plum "is as much like an ordinary person as is possible given this history" (Pereboom 2001). Since it is compatible with this history that Plum have, for example, the ordinary sort of control we have over our environment, the assumption that he lacks this control isn't justified. More generally, this specification rules out Plum's psychology being unusual when it is possible, given the conditions of the set-up, that it is not (cf. Pereboom, forthcoming).

My claim is that there is no reason to suppose that in Case 1 the manipulators cannot, from moment to moment, induce those very neural states had by an ordinary agent, who develops over time in an ordinary ethically reflective and reasons-responsive way. As a result, Plum's memories about past considerations will inform and causally influence his current deliberations. He will also be causally linked to the external world in the proper way. So, for example, "if a bus is careening out of control ready to hop up on the sidewalk and crush him, he is able to respond to those facts and leap from danger" (McKenna, forthcoming). Consequently, if one's intuition is that Plum in Case 1 is not morally responsible, we will have a counterexample to the most prominent compatibilisms, since it won't be possible to avoid this conclusion by claiming that Plum fails to satisfy some more basic condition on agency.

Consider now Case 2, a scenario in which Plum is deterministically programmed by the neuroscientists from the beginning of his life so that his character develops as it does, to all appearances in an ordinary way, as a result of which he is causally determined to kill White while satisfying the prominent compatibilist conditions on moral responsibility. Again, one might

imagine that his mental states and actions exhibit the psychological regularities and counterfactual dependencies of an ordinary agent. Mele claims that in Case 2 how Plum deliberates will not evolve in the normal way, in particular, it will not evolve under the control of his ability to reason critically. A normal agent "might gradually become significantly less egoistic, and, along the way, his less egoistic actions might have reinforcing consequences that help to produce in him increased concern for the welfare of those around him. This increased concern would presumably have an effect on his evolving deliberative habits." But "Plum is cut off from such evolution regarding his procedure for weighing reasons. If anything properly generates the judgment that Plum is not responsible for the killings in [the examples like Case 2], it is this point, for it is the only relevant threat to responsibility that all four cases in the series have in common." (2005, p. 142). Baker makes a similar claim about Plum in Case 2: "He does not kill Ms. White because he wants to will it, but rather because he is executing a nonevolving program" (Baker 2006, 320). [11] However, my description of Case 2 provides no ground for

[11] Baker says, for instance, "Plum is no more reasoning than the paralyzed patient is moving her arm when the arm is being lifted by a physical therapy machine." (Baker 2006, 320)
In Baker's Frankfurt-inspired compatibilist account, a person must be able to conceive of her desires as her own, from the first-person point of view, if she is to be morally responsible. For example, in order for me to be morally responsible for exceeding the speed limit, I must want that I myself will to exceed the speed limit. Here is the complete account (where 'she wants that she* wills X' indicates that she is thinking of herself from the first-person point of view):

(RE) A person S is morally responsible for a choice or action X if X occurs and
(i) S wills X
(ii) S wants that she* will X
(iii) S wills X because she* wants to will X, and
(iv) S would still have wanted to will X even if she had known the provenance of her* wanting to will X.

Now an agent can be manipulated either locally or remotely to have all of these attitudes. In particular, an agent can be so manipulated to occupy the first-person point of view when she is endorsing her first-order volitions, and she can be manipulated so that when she comes to know the provenance of her wanting to will an action, she continues to want to will it. I still have the strong intuition that the agent is not morally responsible under these conditions.

supposing that the evolution of Plum's deliberative style cannot be just as it is for an ordinary agent who develops over time in an ordinary ethically reflective and reasons-responsive way. (Is the thought that a normal deliberative style cannot be produced by the manipulators because this style is outside of the natural causal order?) An agent can be manipulated to possess the very same types of neural states that an ordinary agent has -- both actually and counterfactually -- while his deliberative style is evolving in the way that Mele specifies; (I make this point against Mele's objection in Pereboom 2005; and in Pereboom 2001, 122, I argue that an agent can be manipulated to satisfy Fischer and Ravizza's historical criterion of moral responsibility, which also involves an agent's mental control of his development over time).[12]

Is Plum in Case 2 morally responsible for the crime? It would seem unprincipled to claim that here, by contrast with the local manipulation example, Plum can now be morally responsible just because the length of time between the programming and the action is great enough. Whether the programming takes place a second or thirty years before the action seems irrelevant to the question of moral responsibility. So non-responsibility generalizes from Case 1 to Case 2.

In Case 3, Plum is causally determined by the rigorous training practices of his community, over which he has no control, to kill White while satisfying the compatibilist conditions. It would seem unprincipled to claim that because the manipulation is more ordinary than in Case 2, Plum is now responsible. So non-responsibility generalizes from Case 2 to Case

[12] How Plum's deliberative style evolves would depend on how his environment unfolds, and one might think that this would not be in the Case 2 manipulators' control. But we might imagine that the manipulators know in advance how the environment in which they intend to place him will unfold, and that with this information, their initial programming of his brain issues in their thorough control over how his character develops. Or we might suppose that the manipulators have local control over Plum's environment (but not over his brain), so that if it threatens to result in character, thought, or action that would conflict with their plans, they can make the environment fall into line (thanks to Michael McKenna for raising this concern; cf. Pereboom 2005, 238).

3.

In Case 4, Plum is an ordinary agent in a causally deterministic universe, and he is causally determined to kill White in a way that meets the compatibilist conditions. There would appear to be no difference between Case 3 and Case 4 that could explain in a principled way why Plum would not be responsible in Case 3, but would be in Case 4 (or would not be responsible in Case 1 or Case 2, but would be in Case 4).

Schematically, the argument has three key features. (1) I intend the first two cases, in particular, to pose a direct challenge to the sufficiency of the various compatibilist conditions. (2) I then argue that it is not possible to draw a principled line between any two adjacent cases that would explain why Plum would not be morally responsible in the first, but would be in the second, largely because all of the prominent compatibilist conditions are satisfied in each – this is the "no difference" part of the argument. (3) Finally, I conclude that the best explanation for why Plum isn't responsible in these four cases is that he is causally determined by factors beyond his control in each. The "no-difference" part of the argument and the best explanation part are linked in this way: because there is no difference between Cases 1 and 2, 2 and 3, and 3 and 4 that can explain in a principled way why Plum would not be responsible in the first of each pair but would be in the second, we are driven to the conclusion that Plum is not responsible in Case 4. The salient common factor in these cases that can plausibly explain why Plum is not responsible is that he is causally determined by factors beyond his control to act as he does. I claim that this is the best explanation for his non-responsibility in each of the cases.

One might venture that manipulation by other agents is the best explanation for non-responsibility in Cases 1-3, and that this explanation does not generalize to Case 4. But one

might imagine an example in which Plum is manipulated to act as he does, either locally or remotely, by a mindless machine that has always existed, or spontaneously forms in space (Pereboom 1995; 2001, pp. 115-6; see also Mele 1995, 168-9). It's intuitive that Plum isn't morally responsible in this machine case either, and moreover, there is no difference that can explain in a principled way why Plum is not responsible in Cases 1 and 2, and would be responsible in the machine case. And, I contend, there is no difference between the machine case and Case 4 that can explain in a principled way why Plum would not be morally responsible in the machine case but would be in Case 4.

8. Al Mele's challenge to the incompatibilist conclusion of the four-case argument.

Mele (2004; 2006, pp. 141-4; 2007) argues that in the earlier cases (1 and 2) Plum's being manipulated is a better explanation for his non-responsibility than his being causally determined. In support, he notes that we also have the intuition that Plum is not morally responsible in certain indeterministic but otherwise similar manipulation cases, and that compatibilists have the intuition that Plum is morally responsible in some deterministic scenarios. He then points out that what is common between the deterministic and indeterministic cases is that Plum is manipulated, and he contends that causal determination not being the best explanation is also supported by the intuition that agents can be morally responsible in ordinary causally deterministic scenarios (which don't involve manipulation or other monkey-business). Thus the better explanation for Plum's non-responsibility in these earlier cases is not his being causally determined, but rather his being manipulated.

However, the best explanation in the four-case argument for Plum's non-responsibility in

Cases 1 and 2 cannot be manipulation by other agents, since if in those cases the manipulators are replaced by machines that randomly form in space and that have the same effect on Plum as the manipulators do, the intuition that Plum is not morally responsible persists (Pereboom 2001, 115-6). This is a key part of the argument, and to overturn it, a stronger case needs to be made that manipulation can be the best explanation for non-responsibility in these cases. One might contend that manipulation either by other agents or by machines is the best explanation for Plum's non-responsibility, but then one would need to specify the difference between machine-manipulation and ordinary causal determination that would explain in a principled and satisfactory way why an agent can be responsible in the ordinary case but not when machine-manipulated. This I think cannot be done.

Furthermore, the fact that we can substitute an indeterministic for a deterministic case and still have the intuition that Plum is not morally responsible (in effect I endorse this idea in Pereboom 2001, pp. 41-54) does not show that determinism isn't the best explanation for non-responsibility in deterministic Cases 1 and 2. Here is an analogy.[13] Imagine that a dam at one end of a reservoir would break if the reservoir were filled with more than one billion gallons of water, because the dam could not withstand the pressure that this volume of water would exert. Suppose the reservoir is in fact filled with more than one billion gallons of water, and the dam breaks. It is natural to say here: "what explains the dam's breaking is the water pressure." However, someone might object: "if the reservoir were filled with more than one billion gallons of oil, it would also have broken. So the water pressure doesn't explain the dam's breaking." To this the correct response would be: some true causal explanations set out the actual sufficient

[13] I provide this example in Pereboom (2007, 169-70), and here I embellish my account by responding to Mele's reply in (Mele 2007, 204).

conditions for an event's occurring, and accordingly the explanation by way of the water pressure is true. At the same time, there is an explanation common to both the water pressure and the oil pressure scenarios that is in a sense deeper than the water pressure explanation: liquid pressure higher than a certain level caused the dam to break. Moreover, the water-pressure explanation doesn't compete with the more general, deeper explanation – they are explanations at different levels of generality (Pereboom 2005, 2007). [14]

This example shows that an explanation that lays out sufficient causal conditions for X can be true even if X would still result if some features of those sufficient causal conditions were altered; and also that in such cases there can be a deeper and more general explanation for X, common to both the actual and the counterfactual situation. Thus even if Plum would still be non-responsible for his act of murder if the determinism in Case 2 were changed to some kind of indeterminism, determinism might still explain his non-responsibility in Case 2, while a more general fact, such as the presence of causal circumstances that preclude responsibility-relevant control, might explain his non-responsibility in both cases. In my broader story, causal determinism precludes moral responsibility for the more general reason that it is a type of causal circumstance that precludes responsibility-relevant control, while at the same time there are other types of causal circumstance that also preclude it. Consequently, my broader account cannot be undercut merely by pointing out that there are cases of non-responsibility in which causal determinism is absent, and manipulation is present, while the manipulation results in a type of causal circumstance that precludes responsibility-relevant control.

Mele has recently proposed the following eminently plausible test for a proposed

[14] I agree with Mele's point (2007, 205) that this fact about non-competition of different levels of explanation is not relevant to our disagreement.

explanation: if phenomenon A has possible explanations B and C, see if Bs without Cs result in A, and if Cs without Bs result in A. If Bs without Cs always result in A, and if Cs without Bs do not always result in As, then C is very plausibly not the explanation of A, while B is a good candidate (Mele 2007, 204). Applying this schema to the manipulation argument, the question is: what explains Plum's non-responsibility, causal determination or manipulation? Manipulation (B) without causal determination (C) results in non-responsibility (A), Mele says, and at least some compatibilists find it intuitive that causal determination (C) without manipulation (B) does not always result in non-responsibility (A), and so it seems that causal determination is plausibly not the explanation for non-responsibility, while manipulation is a good candidate.

Mele first has us consider incompatibilists' intuitions about three kinds of cases: "Pereboom's stories featuring manipulation and determinism, my parallel stories featuring very similar manipulation without determinism, and deterministic stories that resemble Pereboom's but involve no manipulation (and no monkey business of other kinds)." Mele then reiterates his claim that because incompatibilists will have the non-responsibility intuition even in his indeterministic versions of the manipulation stories, we are not entitled to conclude that the best explanation for their intuition that Plum is not morally responsible in these examples is that his action results from a deterministic causal process that traces back to factors beyond his control. However,

> …these imagined data do help in the case of compatibilists who have "the intuition" that Plum is not morally responsible for the killing in one or more of Pereboom's stories. Other things being equal, given that that intuition is generated in them both by a

> deterministic case of manipulation and by an indeterministic analogue of that case, but not by a comparable deterministic case involving no manipulation, the manipulation featured in the relevant cases is a better candidate for an explanation of their "nonresponsibility" intuitions about these cases than determinism is. The judgment that the determinism in a deterministic manipulation case provides the best explanation of these compatibilists' "nonresponsibility" intuitions about it is silent on the analogous indeterministic case, and it yields the prediction that these compatibilists will have "the intuition" that Plum is not morally responsible for the killing in any straightforward deterministic story I might tell that involves no manipulation and no monkey business of any kind. Obviously, the imagined data do not warrant that prediction. (Mele 2007, 204-5)

First, as we have already seen, the claim that manipulation without causal determination can issue in non-responsibility is wholly compatible with the hypothesis that causal determinism explains non-responsibility in examples in which the manipulation is causally deterministic. Furthermore, Mele himself agrees that manipulation without causal determination does not always result in non-responsibility. For example, an undetermined agent might be manipulated by an advertisement, or by neuroscientists, to develop a desire for Coke, but since the desire is resistible, and his faculty of practical reasoning remains intact, he is nevertheless morally responsible for deciding to drink one. Manipulation must be of the sort that precludes responsibility-conferring control if it is to result in non-responsibility, and so in order to explain non-responsibility, manipulation needs to be relevantly control-precluding. Part of what's at issue, then, is what it is about Plum's situation that is control-precluding, and it can't be

manipulation per se. Mele conjectures that in my cases the manipulators bypass the capacities for control agents have over their mental lives over time (e.g., 2006, 166-7), and that this is what explains the intuition of non-responsibility. But my exposition of the argument does not support Mele's reading, and in response to objections I have specified how manipulation in my cases would not bypass agent's capacities for cognitively controlled development over time (2001, 122; 2005, 240), and how it would not bypass specifically the sort of mental control that Mele has in mind (2005, 238).[15] Yet the intuition of non-responsibility persists. So this bypass hypothesis won't explain the intuition of non-responsibility in the Plum cases, and my explanatory hypothesis remains in play.

Next, consider Mele's supposition that our assessment of the upshot of the manipulation cases should appeal to the fact at least some compatibilists find it intuitive that in ordinary cases causally determined agents who are not manipulated are morally responsible. To this we can add that many uncommitted participants in the debate, and many incompatibilists, have this intuition as well (Nahmias, Morris, Nadelhoffer, and Turner, forthcoming; Nichols and Knobe, forthcoming). For instance, the incompatibilist Peter van Inwagen would endorse the claim that agents are morally responsible and reject incompatibilism if forced to make a choice (van Inwagen, 2000). And although I would not make van Inwagen's choice, in some cases I find it intuitive that people are morally responsible even under the supposition that they are causally determined, and my intuition that that they are not morally responsible is conflicted. However, the concern that incompatibilists have about these intuitions is that in ordinary cases we are not aware of the actual causes of our actions, and if we were, we would or should reconsider our

[15] In developing his "zygote" manipulation argument, Mele himself recognizes that manipulation need not bypass this sort of control (2006, 184-95).

judgments that agents are free in the sense required for moral responsibility. Spinoza observed, "men think themselves free, because they are conscious of their volitions and their appetite, and do not think, even in their dreams, of the causes by which they are disposed to wanting and willing, because they are ignorant of [those causes]" (Spinoza, 1677/1985, 440).[16] One serious possibility is that our choices and actions result from deterministic causal processes that trace back to factors beyond their control. But our ordinary intuitions and judgments about moral responsibility do not presuppose determinism about choice and action, and they may even presuppose that it is false. Moreover, it stands to reason that these ordinary intuitions and judgments, especially given that they are often emotionally laden, would be resilient even under the supposition of causal determinism (Nichols and Knobe, forthcoming). What's needed is a vehicle for making the supposition of determinism salient in a way that effectively brings it to bear on these intuitions, judgments, and associated emotions. This is part of the point of the examples in the four-case argument. The idea is to devise examples in which the causes of actions are readily salient in this respect, and to show that there is no relevant difference between these causes and ordinary deterministic ones.

It might be claimed that our initial intuition that agents can be morally responsible in ordinary deterministic situations should nonetheless be accorded unmodified weight in our reflective assessment of the manipulation cases. However, to contend that in this assessment the weight of intuitions about cases in which the deterministic nature of the causes has not been made salient should remain unmodified stands in danger of failing to engage the

[16] Another quotation from Spinoza: "experience itself, no less than reason, teaches that men believe themselves free because they are conscious of their own actions, and ignorant of the causes by which they are determined..."(Spinoza 1677/1985: 496).

incompatibilist's challenge. If this challenge is to be engaged, our intuition that in ordinary deterministic cases agents can be morally responsible should not be accorded unmodified weight unless it can be shown that there is a relevant difference between the manipulation cases and ordinary deterministic ones (the soft-line reply, in McKenna's terminology (McKenna, forthcoming)), or supposing that this reply is unavailable, that it is more plausible to judge that Plum is morally responsible in the manipulation examples than that he is not responsible in the ordinary deterministic cases (the hard-line reply). But to my mind, neither of these replies has been established as compelling; (see (Pereboom 2005 and Pereboom forthcoming) for a response to hard-line replies, McKenna's in particular.)

Thus, in answer to Mele, manipulation per se cannot explain Plum's non-responsibility in Cases 1 and 2, nor can manipulation that bypasses agent's ordinary capacities to control their mental lives over time. Causal determination remains a candidate for the best explanation of his non-responsibility in these cases, and this does not conflict with manipulation absent causal determinism yielding non-responsibility in other examples. Finally, if the incompatibilist's concern is to be engaged, our initial intuition that agents can be morally responsible in ordinary deterministic situations should not automatically be accorded unmodified weight in our reflective assessment of the manipulation cases, and so ruling out causal determinism as the best explanation of Plum's non-responsibility by means of this intuition would seem dialectically inappropriate.

9. George Sher on blame and the palatability of hard incompatibilism.

In his recent book, *In Praise of Blame*, George Sher takes my hard incompatibilist

position to task for disallowing blame and allowing only etiolated responses to wrongdoing (Sher 2006). He begins by citing two passages from my article "Determinism as Dente" (Pereboom 1995):

> 1Instead of blaming people, the determinist might appeal to the practices of moral admonishment and encouragement. One might, for example, explain to an offender that what he did was wrong, and then encourage him to refrain from performing similar actions in the future. One need not, in addition, blame him for what he has done. The hard determinist can maintain that by admonishing and encouraging a wrongdoer one might communicate a sense of what is right, and a respect for persons, and that these attitudes can lead to salutary change.

Next, in setting out the view I advocate, I say that the determinist

> 1would resist anger, blame, and resentment, but she would not be exempt from pain and unhappiness upon being wronged. She might, if wronged, admonish, disregard the wrongdoing, or terminate the relationship. (Sher 2006, 5)

In response, Sher writes:

> It is, first, no accident that Pereboom's tone is one of bland, dissociated innocence; for his world without blame, though not without wrongdoing, is devoid of true – that is, freely chosen – evil. Not coincidentally, it is also a world in which human motivation does not conform to familiar patterns. In our actual experience, there is not only nothing to be gained by solemnly informing a carjacker or rapist that he has violated the moral law and really ought to mend his ways, but also little reason to believe that mere hortatory reminders to offenders of what they already know ever have much effect on their

> behavior. In Pereboom's world without blame, by contrast, the prospects for successful jawboning are evidently brighter. In that pastel world, friendly persuasion works. (2006, 5-6).

But first, I argue in *Living Without Free Will*, Chapter 6, that given hard incompatibilism, methods for dealing with criminals other than exhortation are legitimate. These include: detention justified on analogy with quarantining carriers of dangerous diseases, behavioral therapy, and cognitive therapy (Pereboom 2001, 174-86). Second, I was assuming in the quoted passages that the practice of blaming presupposes that its target is morally responsible in the "basic desert" sense: that he deserves to be the object of certain attitudes just because of what he has done -- given at least some understanding of the wrongness of his action -- and not by way of consequentialist considerations. In Sher's own analysis, blame is at its core a certain belief-desire pair; the belief that the agent has acted badly or is a bad person; and the desire that he not have performed his bad act or not have his current bad character (2006, 112). But hard incompatibilism can endorse these beliefs and desires about badness. One might think that if we gave up the belief that people are blameworthy, we could no longer legitimately judge any actions as good or bad. This is mistaken. Even if we came to believe that some perpetrator of genocide was not morally responsible because of some degenerative brain disease he had, we would still very reasonably maintain that it was extremely bad that he acted as he did. Thus in general, denying blameworthiness would not at the same time threaten judgments of badness or desires involving this notion, and, likewise, denying praiseworthiness would not undermine assessments goodness. So far, then, the hard incompatibilist can accept the legitimacy of blame on Sher's account of this notion.

But Sher argues that in addition, blame involves a set of affective and behavioral dispositions, and here one might think his account to conflict with hard incompatibilism. However, he does not regard any of these dispositions as essential to blame, but only connected to it in a looser sense. Given the looseness of this connection, the hard incompatibilist can fully endorse blame in Sher's sense. She might not be able to endorse all of the affective and behavioral dispositions – in particular, not those that presuppose or can only be justified in virtue of basic desert. Still, two important members of this set -- "to apologize for our own transgressions and vices and to reprimand others for theirs" (Sher 2006, 108) – can be accepted, and, as the above quotations indicate, I am explicitly amenable to doing so.

Sher also claims:

> The deepest oddity about Pereboom's world, however, lies not in any of this, but rather in the fact that the only problems wrongdoing appears to present to its inhabitants are future oriented. That, at any rate, is the clear implication of the three responses to wrongdoing – admonish, ignore, walk away – that Pereboom is willing to countenance; for all three recommend themselves primarily as methods of preserving our future tranquility. This exclusively future-oriented stance toward wrongdoing, reminiscent of some of what Strawson says about the objective attitude, is bound to seem profoundly strange to anyone to whom the primary significance of wrongdoing lies not in what it augurs but simply what it is. (Sher 2006, 6)

This isn't right, since there are backward-looking attitudes toward wrongdoing that do not presuppose basic desert. These include being sad about the wrongdoing of another, and regretting one's own wrongdoing (Pereboom 2001, 199-207). The central and essential elements

of blame on Sher's account, which are backward-looking -- the belief that the agent has acted badly, or is a bad person, and the desire that he not have performed his bad act, or not have his current bad character – are not even threatened by any feature of hard incompatibilism.

This discussion of blame is in certain respects representative of the larger upshot of hard incompatibilism. It seems at first that its consequences for human life are seriously impractical, since living in accord with them would undermine far too much of what we value. Closer examination reveals that this is not so. The beliefs and practices at issue survive mostly intact, and relinquishing the features that are undercut would not be as costly as initial appearances would have it.

References

Baker, L. (2006). "Moral Responsibility Without Libertarianism," *Noûs* 40, pp. 307-30.

Campbell, J. (2006). "Farewell to Direct Source Incompatibilism," *Acta Analytica* 21, pp. 36-49.

Fischer, J. M., and M. Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Fischer, J. M. (1999). "Recent Work on Moral Responsibility, *Ethics* 110, pp. 93-139.

Fischer, J. M. (2004). "Responsibility and Manipulation," *The Journal of Ethics* 8, pp. 145-77.

Fischer, J. M. (2006). *My Way*, Oxford: Oxford University Press.

Fischer, J. M., R. Kane, D. Pereboom, and M. Vargas (2007). *Four Views on Free Will*, Oxford: Blackwell Publishers.

Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, pp. 829-39.

Ginet, C. (1996). "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Arguments Convincing," *Philosophical Perspectives* 10, pp. 403-17.

Ginet, C. (2000). "The Epistemic Requirements for Moral Responsibility," *Philosophical Perspectives* 14, pp. 266-77.

Ginet, C. (2002). "Review of *Living Without Free Will*," *Journal of Ethics* 6, pp. 305-9.

Goetz, S. (2002). "Alternative Frankfurt-Style Counterexamples to the Principle of Alternative Possibilities," *Pacific Philosophical Quarterly* 83, pp. 131-47.

Goetz, S. (2005). "Frankfurt-Style Arguments and Begging the Question," *Midwest Studies in Philosophy* 29, pp. 83-105.

Haji, I. (1998). *Moral Accountability*, New York: Oxford University Press.

Haji, I, and M. McKenna (2004). “Dialectical Delicacies in the Debate about Freedom and Alternative Possibilities,” *Journal of Philosophy* 101, pp. 299-314.

Haji, I, and M. McKenna (2006). “Defending Frankfurt’s Argument in Deterministic Contexts: A Reply to Palmer,” *Journal of Philosophy* 103, pp. 363-72.

Hume, D. (1739/1978). *A Treatise of Human Nature*, L. A. Selby-Bigge, ed., Oxford: Oxford University Press.

Hunt, D. (2000). "Moral Responsibility and Unavoidable Action,” *Philosophical Studies* 97, pp. 195-227

Hunt, D. (2005). Moral Responsibility and Buffered Alternatives, *Midwest Studies in Philosophy* 29, pp. 126-45.

Kane, R. (1996). *The Significance of Free Will*, Oxford: Oxford University Press.

Kane, R. (2000). “The Dual Regress Argument and the Role of Alternative Possibilities," *Philosophical Perspectives* 14, pp. 57-79.

McKenna, M. (2003). “Robustness, Control, and the Demand for Morally Significant Alternatives,” in *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities*, M. McKenna and D. Widerker, eds., Aldershot: Ashgate, pp. 201-17.

McKenna, M. (2005). “Where Frankfurt and Strawson Meet,” *Midwest Studies in Philosophy* 29, pp. 163-80.

McKenna, M. (forthcoming). “A Hard-line Reply to Pereboom’s Four-case Argument,” *Philosophy and Phenomenological Research*.

Mele, A. (1995). *Autonomous Agents*, Oxford: Oxford University Press.

Mele, A, and D. Robb (1998). "Rescuing Frankfurt-Style Cases," *The Philosophical Review* 107 (1998), pp. 97-112.

Mele, A. (2005), "A Critique of Pereboom's 'Four-case' Argument for Incompatibilism," *Analysis* 65, pp. 75-80.

Mele, A. (2006) *Free Will and Luck*, Oxford: Oxford University Press.

Mele, A. (2007). "*Free Will and Luck*: Replies to my critics," *Philosophical Explorations* 10, pp. 195-210.

Otsuka, M. (1998). "Incompatibilism and the Avoidability of Blame," *Ethics* 108, pp. 685-701.

Nahmias, E., S. Morris, T. Nadelhoffer and J. Turner (forthcoming). "Is Incompatibilism Intuitive?," *Philosophy and Phenomenological Research*.

Nichols, S. and J. Knobe (forthcoming). "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Noûs*.

Pereboom, D. (1995). "Determinism *Al Dente*," *Noûs* 29, pp. 21-45

Pereboom, D. (2000). "Alternative Possibilities and Causal Histories," *Philosophical Perspectives* 14, pp. 119-37

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

Pereboom, D. (2003). "Source Incompatibilism and Alternative Possibilities," in *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities*, M. McKenna and D. Widerker, eds., Aldershot, U.K.: Ashgate Press, 2003, pp. 185-99.

Pereboom, D. (2005). "Defending Hard Incompatibilism," *Midwest Studies* 29, pp. 228-47.

Pereboom, D. (2006). "Reasons Responsiveness, Alternative Possibilities, and Manipulation Arguments Against Compatibilism; Reflections on John Martin Fischer's *My Way,"*

Philosophical Books 47 (2006), pp. 198-212.

Pereboom, D. (2007). "On Alfred Mele's *Free Will and Luck*," *Philosophical Explorations* 10, pp. 163-172.

Pereboom, D. (forthcoming). "The Hard-line reply to the Multiple-Case Manipulation Argument," *Philosophy and Phenomenological Research*.

Sellars, W. (1963). "Philosophy and the Scientific Image of Man," in *Science, Perception, and Reality*, London: Routledge and Kegan Paul.

Shabo, S. (forthcoming). "Uncompromising Source Incompatibilism," *Philosophy and Phenomenological Research*.

Sher, G. (2006). *In Praise of Blame*, Oxford: Oxford University Press.

Sommers, T. (forthcoming). "More Work for Hard Incompatibilists."

Spinoza, B. (1677/1985). *Ethics*, Appendix to Part I, II 78; *The Collected Works of Spinoza*, ed. and tr. E. Curley, Volume 1; Princeton: Princeton University Press.

Van Inwagen, P. (2000). "Free Will Remains a Mystery," *Philosophical Perspectives* 14, pp.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*, Cambridge: Harvard University Press.

Widerker, D. (1995). "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104, pp. 247-61.

Widerker, D. (2000). "Frankfurt's Attack on Alternative Possibilities," *Philosophical Perspectives* 14, pp. 181-201.

Widerker, D. (2006). "Libertarianism and the Philosophical Significance of Frankfurt Scenarios," *Journal of Philosophy 103*, pp. 163-87.