Defending Hard Incompatibilism

Derk Pereboom, Cornell University

_Midwest Studies_ 29, 2005, pp. 228-47

Penultimate Version

In _Living Without Free Will_, I develop and argue for a view according to which our being morally responsible would be ruled out if determinism were true, and also if indeterminism were true and the causes of our actions were exclusively events. Absent agent causation, indeterministic causal histories are as threatening to moral responsibility as deterministic histories are, and a generalization argument from manipulation cases shows that deterministic histories indeed undermine moral responsibility. Agent causation has not been ruled out as a coherent possibility, but it is not credible given our best physical theories. Hence we must take seriously the prospect that we are not free in the sense required for moral responsibility. I call the resulting view _hard incompatibilism_. Furthermore, contrary to widespread belief, a conception of life without free would not at all be devastating to morality or to our sense of meaning in life, and in certain respects it may even be beneficial.

In addition, I reject a type of incompatibilism according to which the availability of alternative possibilities is crucial to explaining moral responsibility, and accept instead a type of incompatibilism that ascribes the more significant role to an action’s causal history. My view is that an agent’s moral responsibility for an action is explained not by the existence of alternative possibilities available to her, but rather by the action’s having a
causal history of a sort that allows the agent to be the source of her action in a specific way. I thus opt for source as opposed to leeway incompatibilism.

In this article, I will address objections that have been raised to three central arguments that support my version of hard incompatibilism: the argument against the need for alternative possibilities, the argument against compatibilism, and the case against agent causation.

1. Against Requiring Robust Alternative Possibilities

Some leeway incompatibilists have contended that a close examination of Frankfurt-style cases actually substantiates their position. For such cases feature some factor that the intervener’s device is set up to detect that could have occurred but does not actually occur in the agent, such as an intention to do otherwise. The possible occurrence of such a factor -- a "flicker of freedom," to use John Martin Fischer's term -- is then held to be the alternative possibility that is required for moral responsibility. Fischer argues, however, that one can construct Frankfurt-style examples in which the intervener’s device detects some factor prior to the formation of the intention. For instance, one might imagine that Jones will decide to kill Smith only if Jones blushes beforehand. Then Jones's failure to blush (by a certain time) might be the alternative possibility that would trigger the intervention that causes him to kill Smith. Supposing that Jones acts without intervention, we might well have the intuition that he is morally responsible, despite the fact that he could not have done or chosen otherwise, even by forming an alternative intention. He could have failed to blush, but as Fischer argues, such a flicker is of no use to
the libertarian, since it is not sufficiently robust to play a part in grounding moral responsibility.\textsuperscript{5}

I agree with Fischer, and here is a first pass at characterizing robustness. The main intuition underlying alternative possibility conditions is that if, for example, an agent is to be blameworthy for an action, it is crucial that she could have done something to avoid this blameworthiness. If the availability of an alternative possibility per se does in fact play a role in explaining an agent’s moral responsibility for an action, it would have to be robust at least in the sense that as a result of securing that alternative possibility instead, the agent would thereby have avoided the responsibility she has for the action she performed -- it would be her securing of that alternative possibility itself that would explain why the agent would have avoided the responsibility. Failing to blush in the above scenario does not meet this criterion of robustness. For if Jones had failed to blush, he would not thereby have avoided responsibility for killing Smith -- it would not be the failure to blush itself that would explain why Jones would not be blameworthy. By typical libertarian intuitions, a robust sort of alternative possibility would at very least involve the agent’s willing to act in such a manner that would have precluded the action for which he is in fact morally responsible.\textsuperscript{6}

Robustness also has an epistemic dimension, and it is important that it be made explicit in the characterization of this notion. Imagine that the only way in which Jones could have voluntarily avoided deciding to kill Smith is by taking a sip from his coffee cup prior to making this decision, and this is only because it was poisoned so that taking a sip would have killed him instantly. Suppose that Jones does not understand that this action
would preclude his deciding to kill, because he has no idea that the coffee is poisoned. In this situation, Jones could have voluntarily behaved in such a manner that would have precluded the action for which he was in fact blameworthy, as a result of which he would have avoided the moral responsibility he actually has. But whether he could have voluntarily taken the sip from the coffee cup, not understanding that it would render him blameless in this way, is irrelevant \textit{qua} alternative possibility to explaining why he is morally responsible for deciding to kill. Despite the fact that Jones could have voluntarily taken a sip from his coffee cup, and doing so would have rendered him not morally responsible for deciding to kill, this alternative possibility is nevertheless insufficiently robust to have an important role in grounding the agent’s moral responsibility.

Accordingly, an alternative possibilities condition more plausibly relevant to explaining an agent’s moral responsibility for an action must capture the notion that she could have willed otherwise in the following more robust sense: she could have willed something such that she understood that by willing it she would thereby have avoided the moral responsibility she actually has for her action. Here, then, is the notion of robustness I favor:

\textit{Robustness}: 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.
It might now seem that any alternative-possibilities condition on moral responsibility can be defeated by a Frankfurt-style argument that employs a non-robust flicker of freedom. But this line of defense for Frankfurt-style arguments has proven to be too quick. For it is challenged by an important objection to these sorts of arguments that was initially raised by Robert Kane and then systematically developed by David Widerker. The general form of the Kane/Widerker objection is this: for any Frankfurt-style case, if causal determinism is assumed, the libertarian will not have and cannot be expected to have the intuition that the agent is morally responsible. If, the other hand, libertarian indeterminism is presupposed, an effective Frankfurt-style scenario cannot be devised, for any such case will fall to a dilemma. In Frankfurt-style cases the actual situation always features a prior sign by which the intervener can know that the agent will perform the action she does, and which signals the fact that intervention is not necessary. If in the proposed case the sign causally determines the action, or if it is associated with something that does so, the intervener’s predictive ability can be explained. But then the libertarian would not have the intuition that the agent is morally responsible. If the relationship between the sign and the action is not causally deterministic in such ways, then the libertarian can claim that the agent could have done otherwise despite the occurrence of the prior sign. Either way, some alternative possibilities condition on moral responsibility emerges unscathed.

In *Living Without Free Will* and in an earlier article, I develop a Frankfurt-style case designed to evade the objections that have been raised to other versions of such cases. Its distinguishing features are these: the cue for intervention -- the flicker of freedom -- 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 cannot be a robust alternative possibility, and the prior sign -- the absence of the cue -- clearly in no sense causally determines the action the agent actually performs. The following is a revised version of the case I set out in the book:

*Tax Evasion (2)*: Joe is considering whether to claim a tax deduction for the substantial local registration fee that he paid when he bought a house. He knows that claiming the deduction is illegal, that he probably won't be caught, and that if he is, he can convincingly plead ignorance. Suppose he has a very powerful but not always overriding desire to advance his self-interest regardless of the cost to others, and no matter whether advancing his self-interest involves illegal activity. Crucially, his psychology is such that the only way that in this situation he could fail to choose to evade taxes is for moral reasons. His psychology is not, for example, such that he could fail to choose to evade taxes for no reason or simply on a whim. In addition, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to these moral reasons. He 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, Joe could, with his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). More generally, Joe is a libertarian free agent. But to ensure that he choose to evade taxes, a neuroscientist now implants a device, which, were it to sense the requisite level of
attentiveness, would electronically stimulate his brain so that he would choose to evade taxes. In actual fact, he does not attain this level of attentiveness, and he chooses to evade taxes while the device remains idle.

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

This example does feature alternative possibilities that are available to the agent -- his achieving higher levels of attentiveness to moral reasons. But these alternative possibilities are not robust. First, note that in ordinary circumstances, without the interveners device in place, it is not the case that by achieving some higher level of attentiveness Joe would have avoided responsibility for choosing to evade taxes. For under these conditions achieving some higher level of attentiveness is compatible with his not refraining from making this decision, or even ever being seriously inclined so to refrain, and choosing to evade taxes instead.

Still, one might argue, given that the interveners 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. 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 Joe does not have an alternative possibility available to him that is robust. First, he does not even believe that if he had achieved the requisite level of attentiveness he would thereby have been precluded from responsibility for deciding to evade taxes. For he
believes that achieving this level of attentiveness is compatible with his never refraining from making this decision, or even being seriously inclined so to refrain, and deciding to evade taxes instead. In addition, Joe does not know enough to understand that voluntarily achieving the requisite attentiveness would 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 not then have been responsible for this choice. Nevertheless, Joe does not understand that the intervention would then take place, or that as a consequence of this intervention he would be precluded from responsibility for choosing to evade taxes. Hence, no robust alternative possibility is available to him. Still, Joe is morally responsible for deciding to evade taxes.

The core of the Kane/Widerker objection is that if the inevitability of the action given the prior sign is grounded in causal determinism, then the libertarian cannot be expected to agree that the agent is morally responsible for the action, but if we eliminate the causal determination then the agent has robust alternative possibilities after all. But, contrary to Stewart Goetz's criticism,¹¹ in an earlier version of the example, as well as in Tax Evasion (2), the inevitability of the action given the prior sign is not grounded in causal determinism, while at the same time no robust alternative possibilities are available to the agent. In this example, the connection between the prior sign and the action is expressed in the following proposition:

If Joe fails to achieve a certain level of attentiveness to moral reasons, then, provided no one intervenes, he will decide to evade taxes.
The inevitability of Joe’s decision is not grounded in causal determinism since the absence of what would trigger the intervention at some particular time, that is, the absence of a certain level of attentiveness to moral reasons by a particular time, or a state indicated by this absence, does not, together with all the other actual facts about the situation, causally determine the decision. To see this, remove the intervener from the scenario -- we can do so safely, for by hypothesis, the intervener exerts no actual causal influence on Joe’s deciding to evade taxes, so removing the intervener won’t have any implications for whether Joe is causally determined to make this decision. There is no relevant time at which refraining from deciding to evade taxes in the future is impossible for Joe, since he can always achieve the right level of attentiveness, whereupon he can freely refrain from deciding to evade taxes -- or else freely decide to evade taxes. Suppose he does in fact decide to evade taxes and he never achieves the right level of attentiveness. Still, one cannot point to a deterministic process that results in his deciding to evade taxes, for it is never determined that he will fail to achieve the right level of attentiveness, and if he did achieve it, he could then refrain from deciding to evade taxes -- or indeed decide to evade taxes instead.

For the example to be successful against leeway incompatibilism it is not required that the intervention cause Joe to decide to evade taxes. The example would also work if the device, say, rendered him unconscious or killed him or rendered him unconscious instead. For as Fischer points out, it is sufficient for the success of such an example that the agent lacks access to any robust alternative possibility, which would also be the case if the device killed him rather than causing him to choose to evade taxes. In this version of the
example it is perhaps even clearer that determinism is false, since at any time Joe can freely do something that brings it about that he is killed instead of deciding to evade taxes. In addition, such altered formulations of the example turn back a criticism of the original version that Carl Ginet has raised.\textsuperscript{12} He prefers to think of moral responsibility as indexed to precise times, and the question he raises about the example is whether it was open to Joe not to make the choice at the precise time he actually made it – call it $t_1$. Ginet argues that if Joe had made the necessary condition of doing otherwise occur at the last possible moment before $t_1$, the device would have made him choose to evade taxes (a choice that Ginet labels B) after $t_1$, since it would take a little time for the mechanism to work, thus generating an alternative possibility. Now in \textit{Living Without Free Will} I dispute Ginet’s claim about how the mechanism would have to work, and I also dispute the notion that responsibility is best indexed to precise times.\textsuperscript{13} In addition, I claim that the alternative possibility Joe has is still not robust.

Ginet says it is: “for had J taken it, he would at $t_1$ have been refraining from a willing – to do B right then – such that by so refraining he would have avoided responsibility for doing B right then and would have been aware that he was avoiding responsibility for doing B right then (that being such an obvious implication of his not doing B right then, of which he of course would have been aware).” Ginet’s contention depends on the claim that given that the mechanism would have made Joe choose to evade taxes after $t_1$, he would have been aware at $t_1$ that he was not making this choice, and he would have understood at $t_1$ that by not choosing to evade taxes right then, he would not have been responsible for making this choice right then. Now for perhaps a clearer case in which the alternative
possibility is not robust, we can alter the example so that in the neuroscientist’s setup, Joe’s making the necessary condition for doing otherwise occur instantaneously renders him unconscious for several minutes. Then, at the next instant, t1, Joe would not have been aware that he was not choosing to evade taxes, and he would not at t1 have understood that by not choosing to evade taxes right then he would have not been responsible for choosing to evade taxes right then. In these circumstances, Joe would not understood that by making the necessary condition for doing otherwise occur at the last moment before t1, he would thereby have avoided responsibility for not choosing to evade taxes at t1, a moment later.

2. Against Compatibilism

Against source incompatibilism Fischer contends 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."14 Michael Della Rocca points out that this is a claim for which Fischer has not argued, and in fact I believe Fischer can be challenged on this point.15 To be sure, one incompatibilist intuition that we seem naturally to have is that if we could in no sense do otherwise, then we could never have refrained from the immoral actions we perform, and thus we cannot legitimately be held blameworthy for them. But another very powerful and common intuition is that if all of our behavior was "in the cards" before we were born, in the sense that things happened before we came to exist that, by way of a deterministic causal process, inevitably result in our behavior, then we cannot legitimately be judged blameworthy for our wrongdoing. By
this intuition, if causal factors existed before a criminal was born that, by way of a
deterministic process, inevitably issue in his act of murder, then he cannot legitimately be
judged blameworthy for his action. If all of our actions had this type of causal history, then
it would seem that we lack the kind of control over our actions that moral responsibility
requires.

I do not believe that in the dialectic of the debate one should expect Fischer or any
compatibilist to be moved much by this incompatibilist intuition alone to abandon this
position. Rather, here the best type of challenge to the compatibilist develops the claim that
causal determination presents in principle no less of a threat to moral responsibility than
does deterministic manipulation. My four-case argument, first of all develops such
manipulation examples, in which the agent meets all of the prominent compatibilist
conditions on moral responsibility, but which elicit the intuition that the agent is not
morally responsible. But further, it sets out three such cases, each progressively more
like a fourth scenario, one that the compatibilist might envision to be realistic, in which the
agent is causally determined to act in a natural way. The challenge for the compatibilist is
to point out a difference between 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 suggestion is that non-responsibility generalizes from at least one of the
earlier cases to the ordinary one.

In each of the four cases, Professor Plum kills Ms. White for the sake of some
personal advantage. His act of murder 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. But the desire on which he acts is
nevertheless not irresistible for him, and in this sense he is not constrained to act.
Moreover, his desire to kill White conforms to his second-order desires in the sense that he
wills to kill and wants to will to kill, and he wills to kill because he wants to will to kill. In
addition, Plum's desires are modified by, and some of them arise from, his rational
consideration of the relevant reasons, and his process of deliberation is moderately
reasons-responsive. He is receptive to the relevant patterns of reasons -- for instance, if he
knew that the bad consequences for himself resulting from his crime would be much more
severe than they are actually likely to be, he would not have murdered White. But
furthermore, he is not completely egoistic, and indeed he retains the general capacity to
grasp, apply, and regulate his behavior by moral reasons. For example, when egoistic
reasons 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. Now, given that causal determinism is
ture, 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 -- it proceeds from moment to moment. One might
imagine that the sophisticated neuroscientists manipulate Plum from moment to moment
by radiotechnology to act in such a way that these compatibilist conditions are met. The
manipulation takes place at the neural level, so that at the mental level of description the
agent does not differ from one who is not manipulated. Hereby the incompatibilist aims to eliciting the intuition that Plum is not morally responsible, thus showing that the prominent
compatibilist conditions are not sufficient for moral responsibility. Even in an ordinary deterministic situation the fact that the agent meets the compatibilist conditions will not be sufficient for moral responsibility. And I claim that there is no difference between Case 1 and Case 4 – the ordinary deterministic scenario in which the agent meets the compatibilist conditions on moral responsibility -- that can explain why an agent is not responsible when locally manipulated but is responsible in Case 4.

Perhaps the best reply on the part of the compatibilist involves developing the claim that a locally manipulated agent in some sense lacks the metaphysical integrity required to be a morally responsible agent, perhaps that a morally responsible agent must be one whose character develops in the ordinary way over a significant period of time and whose actions proceed from this character and not from external local deterministic manipulation. But then we might imagine 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, and so that he inevitably makes a decision that meets the prominent compatibilist criteria for moral responsibility. Is Plum blameworthy for the crime? The incompatibilist again aims to elicit the intuition that he is not morally responsible. 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 two seconds or thirty years before the action seems irrelevant to the question of moral responsibility. Here again the incompatibilist claims that the absence
of responsibility generalizes to Case 4. (Case 3 is one in which Plum is causally determined by the training practices of his community to have the relevant sort of agency.)

Al Mele offers the following rejoinder to the argument:

Is Pereboom’s judgment about “causal determination” – that is, deterministic causation – warranted? Consider an analogue of case 1 in which the manipulators fall short of initiating deterministic causal chains. In case 1a, the neuroscientists locally manipulate Plum and produce “his every state from moment to moment,” but they do this by means of an indeterministic mechanism. There is an extremely good chance that each push of the button will have the result the neuroscientists want, but each push also has a tiny chance of incapacitating Plum. These are the only possible outcomes of the button pushes. As it happens, Plum is not incapacitated. Here too Plum clearly is not morally responsible for the killing, and the correct explanation of that plainly does not appeal to deterministic causation. That the causation in case 1 is deterministic is not essential to Plum’s lacking moral responsibility in it... Consider an indeterministic analogue of case 2. In case 2a, the program the scientists have installed in Plum is indeterministic. It works just like the program in case 1 except that there is a tiny chance every minute or so that the program will incapacitate Plum. As it happens, Plum is not incapacitated. If Plum is not morally responsible for the killing in case 2, he is not morally responsible for it in case 2a either. Here again the deterministic aspect of the case is not essential to Plum’s lacking moral responsibility for the killing... Pereboom has failed to justify his diagnosis of the alleged source of the intuition that Plum is not morally responsible
for killing White in his cases of manipulation: the deterministic aspect of those cases is dispensable, as I have explained.\textsuperscript{17}

Now I agree that the deterministic aspect of the cases is not essential to Plum's lacking moral responsibility in them. In fact, in \textit{Living Without Free Will} I present an indeterministic manipulation case intended to show that certain kinds of undetermined agents will not be morally responsible. In my view, in either sort of case, deterministic or indeterministic, the agent lacks the sort of control required for moral responsibility. These two sorts of manipulation arguments show that one can lack this sort of control by virtue of being causally determined, and one can lack it by virtue of being a certain sort of indeterministic agent. I also agree that there are many more ways in which one can lack the sort of control at issue. The point of the four-case argument is that determination by factors beyond one's control is sufficient for non-responsibility, for the reason that it precludes the kind of control required for moral responsibility. This point is consistent with the claim that there are other conditions, potentially the theme of other manipulation cases, that are also sufficient for non-responsibility. So determinism's not being essential to Plum's lacking moral responsibility does not undermine the argument.

Mele proposes an alternative diagnoses for our intuitions about non-responsibility in these various sorts of cases. Consider his analysis of the examples like Case 2, in which the agent is not locally manipulated, but rather determined at the outset to behave in certain ways. Mele claims:

\begin{quote}
In each case in this series, Plum played no role at all in shaping his procedure for weighing reasons (say, through trial and error over the years he has been in the
business of deliberating). Unlike normal agents, Plum had no control throughout his history as an agent over this important aspect of his deliberative style. Rather, the procedure is imposed on him – either deterministically or indeterministically – by the neuroscientists or the electromagnetic field. Facts such as these about what Plum does not control in the two series of cases do seem to account for his non-responsibility in them... Normal agents learn how to weigh reasons for action. For example, a young agent who weighs reasons very egoistically may suffer as a consequence and learn that things go better for him when he weighs the interests of others more heavily as reasons. His deliberative style 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. The story of the normal evolution of a particular agent’s deliberative style is a long one. The point here is that in the [Case] 2 series 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 2 series, it is this point, for it is the only relevant threat to responsibility that all four cases in the series have in common... 18

However, one should suppose that the evolution of Plum’s deliberative style in these cases is just like it is in Case 4 -- the ordinary deterministic one -- except for the difference in remote causes.19 One might imagine that Plum in Case 2 is manipulated to possess the very same types of neural states as some ordinary agent has -- even both actually and
counterfactually -- while his deliberative style is evolving. Then, for example, it need not be
that Plum acts from values and principles that are unsheddable because they way they
were programmed bypassed his ability to reason critically about them. Rather, Plum would
be so programmed that he, just as in Case 4, reasons critically about them, and that this
critical process – and it can be as elaborate as one would like – results in his endorsement
of his largely egoistic principles and values. So Plum’s deliberative style in Case 4 evolves
deterministically in a natural way in accord with Mele’s conditions on moral responsibility,
and in Case 2 it evolves deterministically in the same way but for the fact that this evolution
is programmed by the manipulators at the beginning of his life.

One might object that the evolution of a deliberative style would depend on the
details of how Plum’s environment unfolds, and how he responds to these details, and that
these factors could not have been programmed in at the outset. In response, we might
suppose 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 results in their complete control over how his character develops. Or else we might
specify that the manipulators have local control over Plum’s environment (but not over his
brain), so that if it threatens to result in a character that would run afoul of their plans, they
can ensure that the environment falls into line.

Now imagine a scenario in which Plum’s deliberative style is determined by the
manipulators to undergo the sort of evolution Mele thinks is required for moral
responsibility, but he is nonetheless causally determined to kill White. Such an example, I
expect, would generate the intuition that Plum is not morally responsible, and Mele’s
compatibilist account cannot explain this fact. What remains is the diagnosis that in both
cases Plum lacks the control required for moral responsibility, and in each of these cases
his lack of control is explained by the fact that he was causally determined to act as he does.

Fischer has also challenged the four-case manipulation argument. In his view, Plum
is morally responsible in Cases 1 and 2. Our intuition that he is not morally responsible
stems from the correct sense that Plum is not blameworthy.

In my view, further conditions need to be added to mere guidance control to get to
blameworthiness; these conditions may have to do with the circumstances under
which one’s values, beliefs, desires, and dispositions were created and sustained,
one’s physical and economic status, and so forth. Professor Plum, it seems to me, is
not blameworthy, even though he is morally responsible. That he is not
blameworthy is a function of the circumstances of the creation of his values,
character, desires, and so forth. But there is not reason to suppose that anything like
such unusual circumstances obtain merely in virtue of the truth of causal
determinism. Thus, I see no impediment to saying that Plum can be blameworthy for
killing Mrs. White in Case 4. Note that there is no difference with respect to the
minimal control conditions for moral responsibility is Cases 1 through 4 – the
threshold is achieved in all the cases. But there are ... wide disparities in the
conditions for blameworthiness.\textsuperscript{21}

Now there could indeed be cases in which an agent is morally responsible without
being blameworthy – when she is praiseworthy for having performed a good or exemplary
action, or neither blameworthy or praiseworthy – when she performs an action that is
morally indifferent. However, is it not part of our intuitive picture that an agent can be morally responsible for an action, the action be morally wrong, and the agent knows it is morally wrong, while nevertheless the agent is not blameworthy.

In fact, in my view, that an agent is blameworthy for an action is entailed by her being morally responsible for it and her knowing that the action was wrong. For by my characterization, for an agent to be morally responsible for an action is for it to belong to the agent 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, and not, for example, by virtue of consequentialist considerations. Given this characterization, I don’t see how Plum could be morally responsible for committing the murder but not be blameworthy for doing so.

True, there are alternative notions of responsibility that allow for Plum to be morally responsible and not blameworthy. For example, Hilary Bok contends that for an agent to be morally responsible is for it to be legitimate to demand of her that she explain how her decisions accord with morality, and that she evaluate critically what her decisions indicate about her moral character.22 Making these demands of agents might be justified by its effectiveness in improving the agent morally -- we humans are indeed susceptible to causal influence by admonition of this kind. An agent might be morally responsible in this sense and not blameworthy. But this is not the notion of moral responsibility at issue in the free will debate. For incompatibilists would not find our being morally responsible in this sense to be even prima facie incompatible with determinism.
Thus, in my view, if Plum is morally responsible in Case 1 and Case 2 in the central sense that is at issue in the debate – which is the sense that I target – then, in my view, it is entailed that he is blameworthy. An intuition that Plum can be morally responsible without being blameworthy might be explained by the possibility that he might be responsible in some other sense while not being blameworthy. But it is not these senses that either I or incompatibilists generally have thought to be at odds with determinism.

Fischer might argue instead that in Cases 1 and 2 Plum has not gone through the historical process required for moral responsible agency, which he and Mark Ravizza call *taking responsibility.* In *Responsibility and Control* Fischer and Ravizza consider cases that attempt to show that someone’s taking responsibility can be electronically manipulated. There they suggest that such an agent does not in fact meet one of the conditions of taking responsibility: he has not formed his view of himself as an agent subject to the reactive attitudes on the basis of the evidence in the appropriate way. But by analogy to how I argued against Mele, one might imagine that Plum in Cases 1 and 2 is manipulated to possess the very same types of neural states that an ordinary causally determined agent has -- even both actually and counterfactually -- while he is in the process of taking responsibility. Then, some of Plum’s states in Cases 1 and 2 will realize his formation of a view of himself based on the evidence in the appropriate way, just as one might imagine they do in Case 4. As in Case 4, Plum’s formation of his view of himself would be appropriately based on the evidence in these first two cases.
Michael McKenna challenges the four-case argument with a different kind of approach, one that involves foregrounding our intuitions about ordinary cases over those that are elicited by manipulation scenarios:

The compatibilist’s best strategy, it seems to me, is not to show how a suitably determined agent differs so very much from a globally manipulated agent. It is rather to show how similar they are. The compatibilist needs to make clear that once the manipulation is so qualified that all an agent’s current time-slice compatibilist-friendly structures are properly installed through a process of manipulation, then the role of the manipulator begins to shrink into the background; we are simply left with a normal person who happened to be brought into existence in a very peculiar manner. Consider Derk Pereboom’s use of global manipulation cases in his defense of incompatibilism. Pereboom wishes to start with manipulation cases, fix upon the hidden causes that seem to corrupt any appearance of responsibility, and then show how such cases are like standard cases of naturally occurring determination. Once the unseen causes of a naturally determined agent are revealed, Pereboom argues, then our reaction to the agent should be like our reaction to the discovery that a seemingly normally functioning agent has been globally manipulated. The compatibilist should meet Pereboom’s challenge with two moves. First she should work in the other direction, from a (possible) naturally determined agent, to a globally manipulated one. Second, she should fix, not upon hidden causes, but upon the sorts of agential properties that typically serve as a basis for ascribing responsibility. Once it is established that actions issuing from a
(possibly) naturally determined agent invite certain sorts of evaluations in terms of responsibility, one can then hold that actions issuing from an appropriately manipulated agent should be evaluated no differently. The nature of the hidden causes, it can thereby be argued, are not relevant to the sort of psychic structure on the basis of which an agent's responsibility is assessed.26

First a preliminary point. Part of the aim of the four-case argument is to foreground in our assessments of moral responsibility the putative hidden causes of our actions, which are not ordinarily evident, and in particular the fact that they are deterministic. In making moral judgments in everyday life, we typically do not assume that agents' choices and actions result from deterministic causal processes that trace back to factors beyond their control. Our ordinary intuitions do not presuppose that determinism is true, and they may even presuppose that it is false. The incompatibilist's claim is that if we did assume determinism and internalize its implications, our intuitions about whether agents are morally responsible might well be different. 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]."27 The sequence of manipulation cases is intended to elicit the intuition that Plum is not morally responsible by making the deterministic nature of these causes salient. To argue that we should take our cue from cases in which the deterministic nature of the causes is not salient would seem to beg the question against the incompatibilist, for that would amount to a refusal to engage the incompatibilist's challenge.
Is it nonetheless possible to exert pressure on the four-case argument by exploiting the fact that in everyday life compatibilist conditions count as sufficient for moral responsibility? McKenna recommends that one “should fix, not upon hidden causes, but upon the sorts of agential properties that typically serve as a basis for ascribing responsibility.” However, the incompatibilist could welcome fixing on those properties, for on her view they will often be necessary conditions for ascribing moral responsibility. But this does not undermine the claim, which is made intuitive by the manipulation examples, that the absence of determinism is also such a necessary condition. Notice that here the incompatibilist would opt for full disclosure: we should fix on both the hidden causes and the agential properties in question. The suggestion that we should focus only the second would appear to be at a dialectical disadvantage.

McKenna’s considered view is not that we should focus solely on the agential properties, but rather that in assessing the four-case argument, one could legitimately draw greater attention to them, and that this will elicit the intuition that Plum is responsible – certainly in Case 4, but even, for example, in Case 2. At the same time, he allows that drawing greater attention to the hidden causes and their deterministic nature could occasion the intuition that Plum is not morally responsible. But given that each of these two strategies is equally legitimate, the result will be a stalemate. In response, I advocate drawing equal attention to the sorts of agential properties that typically serve as a basis for ascribing responsibility, and to the hidden causes and their deterministic nature by way of the four cases, and then let the intuitions fall where they may. (In fact, in my development of these cases the greater part of each description is devoted to setting out these agential
properties.) Nonetheless, upon following this recommendation, I still have a strong intuition that Plum in Case 4 (and certainly in Case 2) is not morally responsible. To be sure, others may have opposing intuitions. My four case argument will not have persuasive force again them, and I have no conclusive argument to show that they are unreasonable in their reactions.

3. Against agent causation

I defend the view that if we were agent causes, we could be morally responsible. Following many others, I contend that the “luck” objection shows that event-causal libertarianism -- the position that indeterminism in the production of a decision by appropriate events is the key requirement for moral responsibility\(^\text{29}\) -- falls to the objection that if all the causes of actions are events, agents would lack the control over decisions that moral responsibility demands.\(^\text{30}\) In the last analysis, they would lack this control because antecedent causal conditions – including the agent-involving events\(^\text{31}\) – would leave it open whether some particular decision will transpire, and given the causal contribution of these antecedent conditions, the agent has no further causal role in determining whether it does. In fact, given the contribution of these antecedent conditions, nothing – in particular, not the agent herself – settles whether the decision transpires. And if, given these antecedent conditions, the agent does not settle whether the decision transpires, she lacks the control required for moral responsibility for the decision. This might be called “the problem of the disappearing agent.”
At this point, the agent-causal libertarian attempts to specify a way in which the agent could have this further causal role. To solve the problem of the disappearing agent, the agent is reintroduced, this time not merely as involved in events, but rather fundamentally as a substance. The best version of agent-causal libertarianism posits an agent who, fundamentally as a substance, given the causal contribution of the events antecedent to a decision, has the further causal power to cause a decision, upon consideration of reasons, and on the basis of certain reasons, without being causally determined to do so. The proposal is that the control absent on the event-causal libertarian view is supplied by the agent as substance by virtue of her having this causal power.

Agent-causal libertarianism, I contend, does not fall to the luck objection. Recently, Al Mele and Ishtayaque Haji have again disputed this contention, on the ground that the luck objection has as much force against the claim that agent causal libertarianism allows for sufficient control for moral responsibility as it does against the view that event-causal libertarianism allows for this sort of control. First, it is indisputable that when an agent A agent-causes decision D at t, then an event of the following type occurs:

G: A’s causing D at t.

As Mele and Haji emphasize, given exactly the same antecedent conditions as those that precede A’s agent-causing D, G might not have occurred. So that G occurred would still seem to be a matter of luck – it would still seem to be a matter that is not sufficiently under the control of the agent, and in particular the agent does not seem to settle which option, D or not-D, is realized.
The agent-causal libertarian will certainly admit that given these antecedent conditions, G might either have occurred or not. But the core issue is whether the agent nevertheless can have the crucial role in determining whether the decision transpires that she cannot have on the event-causal libertarian’s view. It seems to me that it has not been ruled out that she can. What the agent-cause does most fundamentally is to cause a decision. At this point, one should note that it is a logical consequence of the agent’s causing a decision that an event of type G occurs. It follows logically from the fact that Ann causes the decision to flip the coin that the event Ann’s causing the decision to flip the coin at t occurs. But it is by agent-causing a decision that the agent brings about the event of type G – as a logical consequence of her causing the decision. What thus explains the occurrence of the event of type G – indeed, already given the causal contribution of the antecedent events – is Ann, as a substance, causing the decision. This contrasts with the event-causal libertarian’s scenario, where, given the causal contribution of the antecedent events, the agent has no further role in determining whether the decision occurs. Then events of the form G, given the causal contribution of the antecedent events, will have no further explanation, and, in particular, no further explanation involving the agent.

I want to emphasize at this point that to answer the luck objection, the causal power that the agent exercises in making a decision must be of a different sort from the causal powers of the antecedent events. For the luck objection shows that causal powers of the events are not the sort that can provide the control needed for moral responsibility. And the agent, by virtue of her agent-causal power, is meant to provide this sort of control. Furthermore, given that the agent-causal power is of a different sort from the causal
powers of the events, when a decision is made, the exercise of the agent-causal power will be token-distinct from the exercise of the causal powers of the events.

Can agent-causal libertarianism be reconciled with what we would expect to occur given our best physical theories? If not, given the fact that we have not encountered physical events that are inconsistent with these expectations, we would have reason to doubt this view. Randolph Clarke has in the past defended the claim that on an indeterministic model of physical causation, agent-causal libertarianism can be reconciled with the physical laws. My argument against this claim is as follows. Consider the series of possible actions each of which has a physical component whose antecedent probability of occurring is approximately 0.32. It would not violate the statistical laws in the sense of being logically incompatible with them if for a large number of instances the physical components in this class were not actually realized close to 32% of the time. Rather, the force of the statistical law is only that for a large number of instances it would be correct to expect physical components in this class to be realized close to 32% of the time. Now, are free choices of the sort advocated by the agent-causal libertarian compatible with what the statistical law leads us to expect about them? If agent-caused free action were compatible with what according to the statistical law is overwhelmingly likely, then for a large enough number of instances the possible actions in our class would have to be freely chosen close to 32% of the time. Then, in the long run, the possible actions whose physical components have an antecedent probability of 0.32 will almost certainly be freely chosen close to 32% of the time. But if the occurrence of these physical components is to be settled by the choices of agent-causes, then their actually being chosen close to 32% of the time would
constitute a coincidence no less wild than the coincidence of possible actions whose physical components have an antecedent probability of about 0.99 being chosen, in the long run, close to 99% of the time. The proposal that agent-caused free choices do not diverge from what the statistical laws predict for the physical components of our actions would run so sharply counter to what we would expect as to make it incredible.

In his recent treatment of these issues, Clarke first remarks that the existence of agent causes acting at the macrolevel is consistent with the microlevel being closed under the ordinary indeterministic laws. So all of the components of free and responsible action can be in place given the this sort of microlevel closure: “If there can be substance causation at all, then it seems that there can be substance causation the propensities of the exercise of which conform with complete nondeterministic microlevel causal laws.”37 I think that Clarke is right about this. It is possible that the agent cause act in accord with probabilistic microlevel laws, for it might just happen that in the long run the exercise of agent-causal powers conforms to the probabilities that the indeterministic microlevel laws would assign in the absence of agent causation.

But would one expect this conformity? Timothy O’Connor objects that if the antecedent events are conceived as shaping the agent-causal power, then it is reasonable to expect the actions of agent-causes to conform to the probabilities conferred by these antecedent events:

Imagine that some conscious reasons-guided systems ‘magnify’ microphysical indeterminacies in such a way that several significantly different outcomes are physically possible. Then further suppose that agent-causal power emerges when
conscious reasons-guided systems achieve a requisite threshold of complexity. Such a power might be shaped by states (such as the agent’s reasons for acting) that embody the magnified quantum indeterminacies, so that agent-causal actions would be expected to reflect the physical probabilities in the long run.\textsuperscript{38}

However, to answer the luck objection, the causal power exercised by the agent must be of a different sort from that of the events that shape the agent-causal power, and on the occasion of a free decision, the exercise of these causal powers must be token-distinct from the exercise of the causal powers of the events. Given this requirement, we would expect the decisions of the agent-cause to diverge, in the long run, from the frequency of choices that would be extremely likely on the basis of the events alone. If we nevertheless found conformity, we would have very good reason to believe that the agent-causal power was not of a different sort from the causal powers of the events after all, and that on the occasion of particular decisions, the exercise of these causal powers was not token-distinct. Or else, this conformity would be a wild coincidence, not to be expected and without an explanation. Accordingly, the shaping that O’Connor has in mind cannot be so radical as to undermine the independence of the agent-causal power from the causal powers of the events, and if it is not, then we would expect the divergences at issue.

Against the charge of wild coincidences, Clarke argues that on his integrated view of agent causation\textsuperscript{39} – in which decisions or intention-formations are probabilistically causes by agent-and-reason-involving events and also by agent-causes – conformity with the laws is actually guaranteed:
But with an integrated account, given the way in which agent causation is nomologically tied to causation by events of certain types, instances of agent causation will have to accord with whatever laws govern causation by those events. If these laws derive ultimately from microlevel laws, then the exercise of agent-causal power will have to accord with these microlevel laws.  

But again, for agent-causal libertarianism to solve the luck problem, the causal power of the agent cause must be of a different sort from that of the agent-and-reason-involving events. If the causal power of the agent-cause were exhausted by the causal powers of these events, then we would indeed expect the activity of agent-causes to conform in the long run to the laws governing the events. But then the agent-causal power would not be of a different sort from the causal powers of the events. However, if the causal power of the agent were of a different sort, then we would expect, in the long run, divergences between patterns in which the agent as substance together with the events cause decisions and patterns in which only the events cause decisions. So either the luck problem remains unsolved, or the view generates wild coincidences, which would make it incredible.

4. A Final Word

I have defended three central arguments that support my version of hard incompatibilism: an argument against the need for alternative possibilities, an argument against compatibilism, and a case against agent causation. The first of these aims to secure a significant feature of my own version of hard incompatibilism, since I endorse source incompatibilism. But one could still be a hard incompatibilist and argue that moral
responsibility does require robust alternative possibilities. So of the three, only an argument against compatibilism and an argument against agent-causal libertarianism are germane in the case for any version of hard incompatibilism. To this list we might also add the luck objection against event-causal libertarianism, which I have not defended in detail here. Although this objection is widely endorsed, it may be reinvigorated.

I suspect that views according to which we lack the free will required for moral responsibility have encountered strong resistance partly due to the perception that they would entail giving up too much that is of importance for us. In the last three chapters of *Living Without Free Will*, I argue that this perception is mistaken. I have not revisited this argument here, but together with the main arguments for hard incompatibilism, I think it shows that overall, this sort of position is much more acceptable than is typically believed.
NOTES


9 Tax Evasion (2) differs from the version presented in “Alternative Possibilities and Causal Histories” (p. 128) and in *Living Without Free Will* (pp. 18-9) mainly in its necessary
condition for the availability of a robust alternative possibility. The change is designed to enhance the example’s psychological realism.


19. I make a similar point about Fischer and Ravizza’s theory in *Living Without Free Will*, p. 122.

20. Thanks to Michael McKenna for suggesting this concern.


23. Someone’s taking responsibility for these springs of action has three ingredients: (a) The individual must see himself as an agent; he must see that his choices and actions are efficacious in the world. This condition includes the claim that the
individual sees that if he were to choose and act differently, different upshots would occur in the world.
(b) The individual must accept that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts.
(c) The individual's view of himself specified in the first two conditions must be based, in an appropriate way, on the evidence.

Fischer and Ravizza argue that by proceeding through these steps it is the mechanisms that produce action, in particular, for which the agent takes responsibility; John Martin Fischer and Mark Ravizza, *Responsibility and Control*, pp. 210-4.


25. One might again raise the challenge whether Plum undergoes the right sort of historical process depend on the details of how his environment unfolds, and how he responds to these details, and that these factors could not have been programmed in at the outset. But this objection can be handled in the same way as it was for Mele.


28. In correspondence.


31. That is, events such as: *Joe at (time) t wanting to avoid punishment.*


34. Ishtiyaque Haji, “Active Control, Agent Causation, and Free Action” forthcoming, *Philosophical Explorations*.


36. Note that this is a class of possible actions. I make this specification because it may be that there is no actual class of this sort that is large enough for the argument to go through.


