Philip Swenson

Compatibilism and Control over the Past: A New Argument Against Compatibilism

Michael Moore's recent book *Mechanical Choices: The Responsibility of the Human Machine* is full of rich, insightful discussion of many important issues related to free will and moral responsibility. I will focus on one particular issue raised by Moore: the question of whether we can have control over the past. Moore defends a compatibilist account of moral responsibility on which there are some possible cases in which agents do have such control. But Moore seeks to avoid positing too much control over the past. I will argue that Moore's attempt is not successful. Furthermore, I will try to show that compatibilists in general face a serious challenge regarding control over the past. In sections 1 and 2 I will explore Moore's views. In section 3 and 4 I will offer a general argument against compatibilism about moral responsibility and determinism.

I Moore's Case for Control over the Past

Moore's book *Mechanical Choices: The Responsibility of the Human Machine* contains an intriguing discussion of non causal control over the past. In Chapter 11, Moore considers whether moral responsibility is compatible with a certain sort of epiphenomenalism. Certain interpretations of neuroscientific findings suggest that our choices or acts of will do not cause our bodily movements.[1] Moore does not accept these interpretations, but takes them on board for the sake of argument. If this sort of epiphenomenalism turns out to be correct, can we still be responsible for what we do?

1 Especially relevant are the Libet experiments. See Benjamin Libet, C. A. Gleason, C. W. Wright, and D. K. Pearl, "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness Potential)," *Brain*, Vol. 106 (1983), pp. 623–642

Moore argues that responsibility can coexist with epiphenomenalism. His argument for this involves defending the claim that in some cases, we can have control over the past. Moore focuses on the intriguing example of the parylzed patriot.

The parylzed patriot lives in a world (perhaps like our own) in which willing is epiphenomenal. Our willing to move does not cause our bodies to move. Rather, there is a common cause of both our willing and our bodily movements. Furthermore, scientists can detect a readiness potential signal that always proceeds both our willing to move and the movement itself.

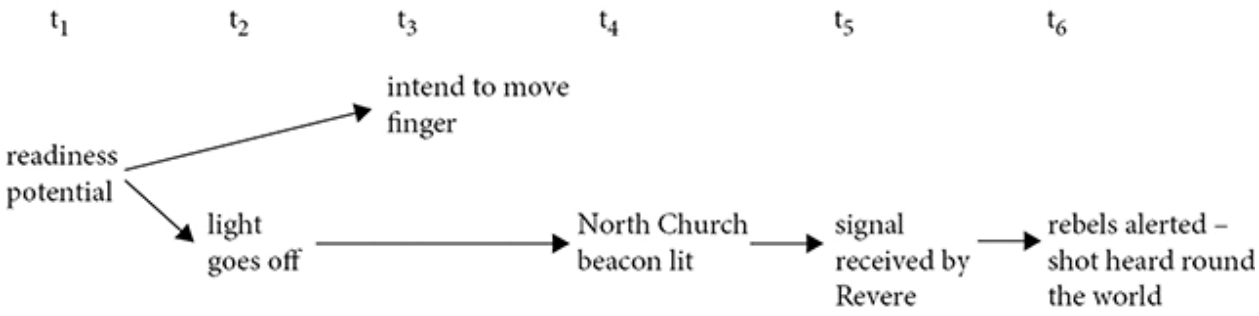
This setup is inspired by a certain interpretation of the results of experiments by Benjamin Libet.[2] Again, Moore does not accept this interpretation, but is considering what we should say if it turns out to be correct.

Here is the case:

> **"the case of the paralyzed patriot (PP)**. As in American history books, PP is in the Old North Church in Boston trying to get the word to the rebels if the British are coming by land or by sea. By prearrangement, one light in the Old North Church tower means by land, two means by sea. ("One if by land, two if by sea," as Longfellow reported it). Yet unlike the actual patriot in the history books (Robert John Newman the sexton of the Old North Church), PP is paralyzed from wounds earlier inflicted by the British. He cannot move his finger to illuminate the lamps. He can will the movement of his finger; but he cannot move his finger because he is paralyzed.
>
> Fortunately PP has been keeping up with twenty-first-century neuroscience. He knows that there are mind-brain interface machines that can read what he is willing and produce some external event to record that fact.[21] Indeed, as it happens, PP is hooked up to one such machine. The machine is attached to the vertex of PP's skull, and it detects slow negative shifts in electrical potential in the SMA [the relevant part of the brain] up to 800 milliseconds prior to PP being aware that he is willing to move his paralyzed finger. The machine's signal is run through an amplifier, and this amplifier signal in turn switches on

2 See Benjamin Libet, C. A. Gleason, C. W. Wright, and D. K. Pearl, "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness Potential)," *Brain*, Vol. 106 (1983), pp. 623–642

> the lights in the tower. Intending to convey the information that he has (that the British are coming by boat), PP wills the movement of his paralyzed finger twice, his SMA emits the distinctive RP [readiness potential] signal twice, and the light goes off twice. The rebels who see the light know the British are crossing the Charles River by boat, and thus know where and when to meet them at Concord. The structure is again that of an epiphenomenal fork, as depicted in Figure 11.5."[3]



Moore's Figure 11.5[4]

Moore thinks PP is clearly responsible for alerting the rebels. He asks rhetorically, "If the British win the war, is not PP fairly hung for treason?"[5] But PP is not responsible because he caused the rebels to be alerted. Rather, PP had non-causal control over whether the readiness potential signal [RP] occurred prior to PP's willing to move his finger. And as a result, he also had control over the known causal results of the occurrence of RP.

Moore maintains that PP's control over the past is based on backtracking counterfactual dependence. He says, "If willing to move one's finger is (backtrackingly) sufficient for RP to have occurred 800 milliseconds earlier, and if the actor knows this and wills the movement of his

[3] Moore, Michael S. (2020). *Mechanical Choices: The Responsibility of the Human Machine*. Oup Usa. Pages 422-423
[4] Moore p. 423
[5] Moore p. 423

finger in order to have had that RP occur, there is control over the known effects of RP as much as if one caused them."[6]

II Too Much Control?

Moore raises a problem for his view. Does accepting that PP has control over the past lead to an implausible amount of control over the past?

He gives another case:

> **Prideful Jim:** "Jim is extremely prideful, so much so that he never asks for favors from his friends if they have recently quarreled and thus given Jim offense. Jim in fact has not asked Jacob for a favor today because they quarreled yesterday. Suppose the following counterfactual is true: if Jim had asked Jacob for a favor today, then it would have to have been the case that they did not quarrel yesterday. Would anyone suppose that Jim's decision today about asking for a favor controls whether Jim and Jacob quarreled yesterday?"[7]

Moore's answer is that Jim's decision today does not control the past. (And I agree.)

But in this case, just as in the case of the parylzed Patriot, there is backtracking counterfactual dependence. As Moore says, "If Jim had asked Jacob for a favor today, then it would have to have been the case that they did not quarrel yesterday."[8]

So the worry for Moore is as follows:

> *If we maintain that PP has control over the past because the past counterfactually depends on his future choice, we will end up with too much control over the past. We may have to say that agents like Jim have control over the past as well.*

---

[6] Moore p. 425

[7] Moore p. 427. Moore also presents two other similar cases. Moore adapts the Prideful Jim case from P. B. Downing, "Subjunctive Conditionals, Time Order, and Causation," *Proceedings of the Aristotelean Society*, Vol. 59 (1959), pp. 125–140

[8] Moore p. 427

Suppose there are a great many true backtracking counterfactuals of the form *if later event e had not occurred then earlier event c would not have occurred.* If so, then there will be many agents like Jim. And, if backtracking counterfactual dependence is sufficient for control over the past, we will have to posit an implausible amount of control over the past.

We could say that you only have control over the past when you know that the relevant backtracking counterfactuals are true. But then we would still get the strange result that if only agents were aware of certain counterfactuals, they would have control over the past.

Moore's suggests a solution to this problem. We should limit control over the past to cases involving *strong necessity.* If I understand him correctly, he would endorse the following principle:[9]

> **strong necessity principle:** An agent has control over the past only when a later event e within the control of the agent strongly necessitates some earlier event c.

What is strong necessitation? For Moore, strong necessitation involves counterfactual dependence continuing to hold in "more remote" possible worlds.

Moore says "I have in these examples relied on an intuitive equation of strength of necessitation with closeness of possible worlds; the more distant are the possible worlds in which the backtracking counterfactual remains true, the stronger the necessitation of the earlier event by the later choice."[10]

So consider a backtracking counterfactuals such as:

B: *If later event e had not occurred then earlier event c would not have occurred.*

I take Moore's view to be something like:

---

[9] See the discussion on Moore p. 432-433
[10] Moore p. 433

if ***B*** remains true even in remote possible worlds, then e strongly necessitates c.

This handles the case of Prideful Jim. Moore says: "There are many possible worlds (and not even very distant ones) where Jim would doubtlessly overcome his pride and ask for the favor he so desperately needs."[11]

However, I do not think the appeal to strong necessity will work. We can imagine a variety of scenarios where PP's willing does not strongly necessitate the readiness potential signal. Consider:

Scenario 1: Suppose that future scientists discover an artificial way to cause readiness potential in the brain. When they use this artificial method, readiness potential does not normally precede intention/willing. The correlation is broken. PP knows that scientists have recently been utilizing this artificial method in some people. But he also knows that the alternative method is not active in his brain. So he knows that if he does not will the movement of his finger, readiness potential would not occur. But there are now fairly nearby worlds in which it is false that *RP would not have occurred if PP had not willed to move his finger* (worlds in which the scientists utilize the artificial method in PP's brain).

Scenario 2: Suppose it turns out that the laws of nature are rather strange, and intentions are only consistently preceded by readiness potential on Wednesdays. On all other days readiness potential occurs randomly and does not correlate with willings. Suppose that PP knows all this, and knows that it is Wednesday. Now there are fairly nearby worlds in which it is false that *RP would not have occurred if PP had not willed to move his finger* (worlds in which it is not Wednesday).

Scenario 3: Suppose there is a powerful angel who is unsure whether to foil PP's plans. He decides to flip a coin. If it lands heads he will perform a miracle and make it the case that readiness potential in PP's brain occurs randomly and does not correlate with willings. However, the coin lands tails and the angel does nothing. (And PP knows all this.) So PP's belief that *RP would not have occurred if PP had not willed to move his finger* is still true. But it is now false in nearby worlds (worlds where the coin landed heads).

It doesn't seem to me that these additions to the story should change our evaluation of PP's responsibility. Since PP still knows about the counterfactual connections between his choices and the past, it seems that he is still responsible. If he was in control over the past in the

[11] Moore p. 433

initial case, then he is still in control in **Scenarios 1-3**. But this conflicts with the **strong necessity principle.**

In my case at least, the intuition that PP is responsible is sustained so long as PP knows both of the following:

-If he had willed to move his finger then the rebels would have been alerted.

And;

-If he had not willed to move his finger then the rebels would not have been alerted.

Since these conditionals can be true in the absence of strong necessity, I do not think the appeal to strong necessity succeeds in solving the problem. Accepting that PP has control over the prior occurrence of RP still threatens to lead to an excessive amount of control over the past.

After reflecting on the case of PP, here are two claims on which I find myself in agreement with Moore.

1. If there really are free agents like PP, it's intuitive that they are responsible in virtue of the connection between their choices and the past.
2. It is implausible to posit a very large amount of control over the past.

I will now try to use these two insights to generate an argument against compatibilism.

III A General Argument Against Compatibilism?

Compatibilists say that agents can be morally responsible even if causal determinism is true**.** I will now try to make trouble for Compatibilism.

Consider this scenario:

> **Case C**: James knows that an eccentric billionaire donated a large sum to famine relief yesterday if and only if P&L is true. (P is a comprehensive purported description of the

> past and L is a comprehensive purported description of the laws of nature.) James also knows that P&L is true if James raises his hand now and false otherwise.
>
> How does he know all this? Suppose he knows that either P&L or P*&L* describe the past and laws. (Perhaps reliable scientists have discovered this and informed him.) And he knows that *if P&L is true then the billionaire is determined to donate and he is determined to raise his hand.* And he also knows that *if P*&L* is true then the billionaire is determined not to donate and he is determined not to raise his hand*.
>
> James does not know whether the billionaire donated the money (or whether P&L is true). Suppose also that raising his hand would come at a very minor cost to James. As it turns out, James does not raise his hand.

It seems to me that, if compatibilism about responsibility and determinism is true, then James can be in **C** while both of the following are true:

> (a) James has the sort of control over raising his hand which is required for moral responsibility.
>
> (b) James raising his hand would not cause (or otherwise explain) past events.

Why does Compatibilism allow for the truth of both (a) and (b)? Consider first (a). James knows that either he is causally determined to raise his hand or he is causally determined not to. But that would be true of any agent who knows that determinism is true. I doubt many compatibilists would want to say that knowing determinism is true is incompatible with moral responsibility. (Incompatibilists will of course deny (a), since **C** involves James' action being determined.)

Importantly, James does not know whether he is going to raise his hand. Even if knowing what one will do ahead of time undermines control, this would not apply to James. It is true that James knows some weird conditionals. But knowing these conditionals does not appear to conflict with standard compatibilist accounts of control. James could still possess the conditional ability to do otherwise which classical compatibilists (such as Moore) regard as necessary for moral responsibility. And James could still be reasons-responsive (as required by many

semi-compatibilists).[12] So I think both classical compatibilists (who say that the ability to do otherwise is compatible with determinism) and semi-compatibilists (who claim that moral responsibility does not require the ability to do otherwise) should grant (a).

Regarding (b), Moore does not think we should posit backward causation in the case of the parylized patriot, and I think the same is true for case **C**. Suppose that both of the following backtracking conditionals are true:

-If James had raised his hand, the billionaire would have donated.

-If James had refrained from raising his hand, the billionaire would not have donated.

One could claim that the truth of these conditionals is sufficient for backward causation. However, the resulting picture seems implausible. Even if determinism is true, it should not automatically turn out that there are a great many cases of backward causation. So we should either deny, as Moore does, that backtracking counterfactual dependence suffices for causation. Or we should deny that these backtracking conditionals are true. (The Local Miracle view discussed below endorses the later claim.)

So I think compatibilists should not say that James being in **C** requires backward causation. It looks as though, given compatibilism, both (a) and (b) can be true.

I now want to advance the following claims:

> **Key Claim 1:** Assuming (a) is true, James is blameworthy for not raising his hand in virtue of the connection between his raising his hand and the past.
>
> **Key Claim 2:** Assuming (b) is true, James cannot be blameworthy in virtue of the connection between his raising his hand and the past. James does not have the right sort of control over the past.

[12] For a semi-compatibilist account of responsibility See Fischer, J. M., & Ravizza, M. (1998). Responsibility and control: a theory of moral responsibility. Cambridge: Cambridge University Press.

Consider **Key Claim 1**. If James is in control of raising his hand but does not do so, he seems blameworthy. He knows that a good outcome is guaranteed if he raises his hand, and a bad outcome is guaranteed if he does not. He seems like a real jerk if he doesn't raise his hand. Furthermore, he seems blameworthy because of the connection between his hand raising and the billionaire's past decision. Absent that connection, not raising his hand would be perfectly fine.

But now consider **Key Claim 2.** If there is no explanatory chain running from James to the billionaire's past choice, how could he be blameworthy in virtue of his connection to this choice? It seems to me that, without backward causation, he cannot have control over the past. And without control over the past, it seems that he should not be blameworthy in virtue of the connection between his hand raising and the billionaire's past decision.

But **Key Claim 1** and **Key Claim 2** cannot both be true. Perhaps the best way to make sense of all this is to say that **C** cannot obtain while (a) and (b) are true. But, as we have seen, compatibilism makes it plausible that that **C** *can* obtain while (a) and (b) are true. This suggests that compatibilism is false.

Here's another way to put the argument. Compatibilism gets us into a situation in which all three of these claims seem true:

> **-If James has no control over the past, then he is not blameworthy in virtue of the connection between his raising his hand and the past.**
>
> **-James has no control over the past (assuming no backward causation).**
>
> **-James is blameworthy in virtue of the connection between his raising his hand and the past.**

But you cannot accept all three. The way out is to reject compatibilism.

The argument thus far has relied on the thought that in the absence of backward causation, there cannot be control over the past. This claim seems very plausible to me, but Moore would reject it, since he thinks PP has non-causal control over the past. What Moore objects to is ubiquitous noncausal control over the past, not the existence of some noncausal control over the past. Perhaps compatibilists could accept that James has non-causal control over the past but deny that such control is problematically common.

However, I worry that accepting that James has control over the past would lead to ubiquitous control over the past. Notice that the only strange thing about case **C** is the sort of knowledge James has. If determinism is true, then there are probably many cases in which both future event e and and past event c are determined by the actual past and laws, but would both be determined not to occur by some rival past and laws. If we add the supposition that the past counterfactually depends on James' choice, that too may be a common occurrence (as we saw with the discussion of Prideful Jim in section 2).

Suppose we think of knowledge and control as independent features of an agent. If so, then, when it comes to control, James' case is not unique. If we accept that James has control over the past, then we should accept that we often have control over the past. It is only our lack of knowledge that keeps us from regularly being responsible for past events. If this is correct, even Moore should be worried.

One could respond by claiming that knowledge is necessary for control. Since normal agents lack the sort of knowledge James has, they lack control over the past. But even so, we would still have the strange result that, *if we only knew more about the past, we would have control over it.*

Objection 1: James is not blameworthy for raising his hand because causal decision theory is true. According to causal decision theory, we only need to consider the causal upshots of our choices when making decisions.[13] Since there is no backward causation in **C**, James is not blameworthy for ignoring the connection between his choice and the billionaire's donation. (PP is not responsible for the same reason.)

Response: causal decision theory is implausible if we can make free choices in deterministic scenarios. James knows that it's *guaranteed* that people will be saved from famine if he raises his hand. It is crazy to ignore this sort of information.

Also, imagine you are in a standard prisoners' dilemma. Defecting is better for you than cooperating no matter what the other player does. But both of you cooperating is significantly better for you than both defecting.

Now suppose that you learn that you and the other player are living causally determined duplicate lives. It is causally determined that he will make all of the same choices as you. Whatever you are causally determined to do, he is determined to do the same. Doesn't it seem clear that it would be crazy to defect in this case? You are absolutely certain to be worse off if you do.[14] Causal decision theory looks bad in deterministic cases.[15]

---

[13] For an overview of causal decision theory see Weirich, Paul, "Causal Decision Theory", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/decision-causal/>.

[14] Thanks to Brian Cutter for the example.

[15] David Lewis also drew a connection between prisoners' dilemmas and decision theory. See Lewis, David K. (1979). Prisoners' dilemma is a newcomb problem. *Philosophy and Public Affairs* 8 (3):235-240.

Objection 2: James is blameworthy, but not in virtue of the connection between his raising his hand and the past. Rather he is blameworthy because he displayed a bad quality of will.

Response: James only displays bad quality of will if his connection to the billionaire's past donation gives him a reason to raise his hand. So his displaying bad quality of will still depends on the connection between his raising his hand and the past.

One might claim that James displays bad quality of will so long as he merely believes (even falsely) that there is a connection between his choice and the billionaire's past donation.

I agree. But the reason he displays a bad quality of will is because, if his beliefs were true, the connection between his choice and the billionaire's past donation would be relevant to what he should do. *If his beliefs were true, the connection between his choice and the billionaire's past donation would give him a reason to raise his hand.* Thus, this account of James's blameworthiness still requires the connection between his raising his hand and the past to play a role. And this is still implausible if James does not possess control over the past.

Objection 3: The past does not counterfactually depend on James's choice, so he isn't blameworthy. It is not true that if he had raised his hand the billionaire would have donated the money.

Local Miracle Compatibilists claim that in deterministic scenarios, it is not true that if you had acted differently then the distant past would have been different. Rather, if you had acted differently a miracle or law breaking event would have occurred shortly before your act.[16] So if

[16] See Lewis, David (1981). Are we free to break the laws? *Theoria* 47 (3):113-21.

James had raised his hand, the billionaire would still not have donated. Instead, a miracle would have occurred right before James raised his hand.

> **Response A:** I am not sure that counterfactual dependence is needed in order for James to be blameworthy. James knows that in every epistemically possible scenario in which he raises his hand, the billionaire also donates. The probability that the billionaire donates, given that James raises his hand, is one. And the probability that the billionaire donates, given that James does not raise his hand, is zero. Given those facts it is still fairly intuitive that James should raise his hand and is blameworthy if he does not.
>
> **Response B:** Suppose the billionaire has a team of physicists who can detect whether a law breaking event (a departure from the laws described by L*) occurs in the vicinity of earth. The billionaire is committed to donating only if some lawbreaking event is detected. And James knows this. James does not know whether L* is true. But he does know that if he raises his hand, a departure from L* would have occurred shortly before. In this case the relevant part of the past does counterfactually depend on what James does now. So the Local Miracle view still involves backtracking counterfactual dependence.

Objection 4: You are trying to have your cake and eat it too. The intuition that James is blameworthy (expressed in **Key Claim 1**) is supported by evidential decision theory. Evidential decision theory maintains that it is rational to perform the action that provides the most evidence of good outcomes (even if your act will not cause those good outcomes). Evidential decision theory implies that the past is relevant to what James should do, since what James does impacts

his evidence that good things happened in the past. So according to evidential decision theory, James is irrational for not raising his hand.[17]

But the intuition that James cannot be blameworthy (expressed in **Key Claim 2**) relies on causal decision theory. Recall that causal decision theory maintains that only potential causal upshots of our choices impact what it is rational to do. So it delivers the result that the past is irrelevant to what James should do. You are relying on incompatible accounts of the correct decision theory to motivate your argument.

> **Response:** I am not relying on the actual truth of evidential decision theory to motivate **Key Claim 1.** I maintain that if James really could make a free choice in C, then he would be blameworthy. So I agree that if James really could make a free choice in C, then evidential decision theory would deliver the correct result. But as an incompatibilist I maintain that James cannot make a free choice in C. So I do not need to say that evidential decision theory is true. My reason for endorsing **Key Claim 1** is intuition, not the actual truth of evidential decision theory.
>
> (In fact, I am inclined to think that evidential decision theory delivers the wrong results in more ordinary sorts of cases. Suppose it turns out that eating spinach only correlates with, but does not cause, good health. People who care a lot about being healthy tend to eat spinach and also tend to be healthier. But there is no causal link. And suppose you knew this for sure. It might still be the case that eating spinach with the aim of becoming healthy would give you evidence that you will be healthy, since it might give you some evidence that you care a lot about health. (After all, you are acting on the

[17] For an overview of evidential decision theory see Ahmed, Arif (2021). *Evidential Decision Theory*. Cambridge University Press.

motive of pursuing health.) Evidential decision theory seems to imply that it would then have reason to eat spinach because you want to be healthy. But that seems silly.)[18]

Additionally, I maintain that causal decision theory delivers bad results if we can make free choices in deterministic scenarios. But I do not think we can make free choices in deterministic scenarios. So causal decision theory may still be correct.

Objection 5: You are still relying on the tension in our intuitions that leads to the conflict between evidential decision theory and causal decision theory. But perhaps that same tension can be drawn out in indeterministic cases. If so, rejecting compatibilism would not solve the problem.

Let's imagine a variation on case **C** that does not involve causal determination of James' choice. Suppose James knows that if P&L is true it's 90% likely that he raises his hand. But if P*&L* is true it's only 10% likely he raises his hand. And he still knows that if P&L is true then the billionaire is determined to donate. And he also knows that if P*&L* is true then the billionaire is determined not to donate. In this variation James can be free and responsible even if incompatibilism is true. And it is still much more likely that the billionaire donated if he raises his hand than if he doesn't raise his hand. Does he still seem blameworthy for not raising his hand?

Response: First, Some incompatibilists deny that an act can be free (in the sense required for moral responsibility) if there is a fixed objective probability that the agent will perform the act.[19] If that is correct, then James cannot be blameworthy in this version of

[18] Weirich 2020 presents a similar sort of ordinary case.
[19] See e.g. Chapter 3 of Pereboom, Derk 2014. Free Will, Agency, and Meaning in Life, Oxford: Oxford University Press.

the case either. However, I myself am not inclined to agree with this view. So I won't put too much weight on it.

Second, I do not have a clear intuition that James is blameworthy in this indeterministic case. In this case James' evidence no longer rules out the possibilities that (i) he doesn't raise his hand and the billionaire still donates or (ii) he raises his hand and the billionaire still does not donate. My sense that the billionaire's possible donation gives James a reason for action dissipates. James might reason along these lines. 'Either P&L is true or P*&L* is true. There's nothing I can do to have an impact on which of them is true or to guarantee that one of them will be true. Furthermore, both are consistent with any action I take. So I might as well do whatever has the best casual upshot.' This sort of reasoning seems bad when it is guaranteed that the billionaire's choice will correlate with James' act. But it is not so clearly bad in this case.

Third, it is plausible that we no longer have backtracking counterfactual dependence of the past on the future. Suppose P*&L* is true, James did not raise his hand and the billionaire did not donate. If James had raised his hand, would P*&L* have been false? And would the billionaire have donated? I'm inclined to think not. Worlds in which something improbable happens (a 10% chance eventuates) seem closer than worlds in which the past is different. So I am inclined to say that if James had raised his hand the billionaire would still have failed to donate. Insofar as the motivation for holding that James is blameworthy in **C** depends on backtracking counterfactual dependence, the motivation does not carry over. (But note the tension between this response and my Response A to Objection 3 above.)

Objection 6: Let's imagine a more extreme version of the case. Suppose James knows that if P&L is true it's 99.999% likely that he raises his hand. But if P*&L* is true it's only 0.001% likely that he raises his hand. He still knows that if P&L is true then the billionaire is determined to donate. And he still knows that if P*&L* is true then the billionaire is determined not to donate.

Here James cannot guarantee the billionaire's donation. But he can make it *extremely* likely. Doesn't it seem strange to hold that James would be blameworthy in **C**, but not in this very similar case?

Response: I agree that it seems strange to think that a move from 100% guaranteed to only 99.999% guaranteed could make a difference. Nonetheless my intuition does shift. The 'I can't make an impact or guarantee a result' reasoning described above still seems more sensible than in the deterministic case **C**. I think the responses given to Objection 5 still have some force here.

In addition to those responses, there is a further move for dealing with this extreme case available to the incompatibilist (available even to the incompatibilist who is not a responsibility skeptic). The incompatibilist is not committed to saying that any level of probability short of 100% is consistent with free agency. Perhaps an act's having an extremely high probability undermines freedom even if the probability is not 100%. This would introduce vagueness into the picture. (Precisely how high of a probability is too high?) But vagueness pops up all over the place. So if the 99.999% version of case **C** generates the same problems as the deterministic case, it is open to the incompatibilist to deny that James can act freely in the 99.999% version of case **C** as well.

IV: Another Version of the Argument

I will now present a version of the argument that does not rely on the claim that James is blameworthy in case C. Here is the argument:

P1 If compatibilism is true, James has control (in the sense required for moral responsibility) over raising his hand.

P2 If James has control over raising his hand, then he has control over the billionaire donating yesterday.

P3 James does not have control over the billionaire donating yesterday.

P4 So, James does not have control over raising his hand.

P5 So, compatibilism is false.

P2 is a key premise. What justifies it? Here is a plausible principle regarding the transfer of control:

Control Transfer (Conditional Version):

If

(1) you have control (in the sense required for moral responsibility) over doing A, and
(2) you know that e occurs if and only if you do A,

then you have control (in the sense required for moral responsibility) over e.[20]

Control Transfer appears to support P2. To satisfy (2), we need the claim that the billionaire donated if and only if James raises his hand. And this seems like an intuitive thing to say about case C (although I acknowledge that indicative conditionals are notoriously tricky).

[20] Thanks to Carolina Sartorio for helping me see that the argument could make use of a transfer principle like Control Transfer.

Here is another transfer principle that supports P2:

> Control Transfer (Probability Version):
>
> If
>
> (1) you have control (in the sense required for moral responsibility) over doing A and
>
> (2)* the epistemic probability of e is 100% conditional on your doing A and 0% conditional on your refraining from doing A
>
> then you have control (in the sense required for moral responsibility) over e.

This principle is attractive so long as epistemic certainty entails truth (i.e. a 100% epistemic probability in P entails P).

So we have two Control Transfer principles that, if true, appear to support P2. And both principles seem intuitively plausible. However, they both face an apparent counterexample. Consider this case:

> **Copycat:** Chris intends to wear a red hat today only if Aaron also wears a red hat. Chris has no control over Aaron's hat choice. But he knows for sure that he will be able to observe whether Aaron wears a red hat and that he will decide to wear a red hat only if he sees Aaron wearing one. [21]

Chris has control over his wearing a red hat. According to the case, he knows for certain that he will wear a red hat if and only if Aaron does, and yet he does not have control over Aaron's hat choice. I myself am not convinced that it is possible to know in advance what you will freely do in the future. But if we grant that such foreknowledge is possible, **Copycat** looks like a clear counterexample to both versions of Control Transfer.

[21] Thanks to Chris Tucker for pointing out this sort of case to me.

What is going on here? One way for an agent's free act A to perfectly correlate with event e is for the agent to control e in virtue of controlling A. **Copycat** reveals a second way such a correlation may arise. The agent may lack control over e, but instead have control over whether e correlates with A. Chris does not have control over Aaron's hat choice, but he has control over whether his choice matches Aaron's. Both versions of Control Transfer fail to account for this possibility

. The insight behind the Control Transfer principles is that there cannot be an event e that (i) you lack control of and that (ii) ***unavoidably*** perfectly correlates with something that you have control over. **Copycat** reveals that having control over the correlation itself is a way for (ii) to fail to be met. Here is a principle that avoids counterexamples like **Copycat**:

Control Transfer III:

> If
>
> (1) you have control (in the sense required for moral responsibility) over doing A, and
> (2) you know that e occurs if and only if you do A, and
> (3) you don't have control over the fact that e occurs if and only if you do A,
>
> then you have control (in the sense required for moral responsibility) over e.[22]

Control Transfer III avoids the **Copycat** counterexample because (3) is not met in **Copycat**. Chris does have control over whether his hat choice matches Aaron's.

Does Control Transfer III still support P2? Recall:

P2 If James has control over raising his hand, then he has control over the billionaire donating yesterday.

---

[22] Control Transfer (Probability Version) could be modified in a similar fashion.

Control Transfer III still supports P2 unless James has control over:

B: James raises his hand if and only if the billionaire donated yesterday.

If James had control over B, then 3 would fail to be met. And Control Transfer III would not support P2. But I don't find it very plausible that James has control over B. It is plausible that B would be true no matter what James does. In the nearest world where James does not raise his hand, it is plausible that P*&L* is true. Other possibilities would require that the scientists who informed him that either P&L or P*&L* be mistaken or have made different predictions. This seems to make these possibilities more distant. So James does not appear to have counterfactual control over B.

Furthermore, if B would be true whether James raises his hand or not, it does not seem plausible that James' act causes B to be true. Of course, there are overdetermination and preemption cases in which causes fail to change the outcome. However, there seems to be no better reason for thinking that James' raising his hand is a cause of B than there is for thinking that his failure to raise his hand would be a cause of B. Carolina Sartorio has proposed this attractive principle:

> DM2-CAUSATION: Causes make a difference to their effects in that the effects would not have been caused by the absence of their causes.[23]

DM2-CAUSATION seems like an attractive test for causation. And since James raising his hand is not a better candidate for causing B than his failure to raise his hand would be, it is hard to see how we get the asymmetry required by DM2-CAUSATION.

---

[23] See Sartorio, C. (2013) 'Making a Difference in a Deterministic World', The Philosophical Review,122:189–214.

If James's act neither makes a difference with regard to B or causes B, it is implausible that James has control over B. And if James does not have control over b, Control Transfer III still supports P2.

I conclude that there is an attractive argument against compatibilism based on the results to which compatibilism leads regarding control over the past.[24]

[24] For helpful comments or discussion thanks to Zach Barnett, Brian Cutter, Kenny Boyce, Dustin Crummett, Doug Husak, Alex Kaiserman, Paul Manata, Austen McDougal, Michael Moore, Carolina Sartorio, Dan Speak, Raymond Blaine Stewart, Neal Tognazzini, Chris Tucker, Kadri Vihvelin and Alec Walen. Thanks also to audiences at FLAW, Rutgers University, Uppsala University and William & Mary.