The Philosophical Quarterly Vol. 0, No. 0 *2020*
ISSN 0031-8094 doi: 10.1093/pq/pqaa061

ON THE COMPATIBILITY OF RATIONAL DELIBERATION AND DETERMINISM: WHY DETERMINISTIC MANIPULATION IS NOT A COUNTEREXAMPLE

By Gregg D. Caruso

This paper aims to defend deliberation-compatibilism against several objections, including a recent counterexample by Yishai Cohen that involves a deliberator who believes that whichever action she performs will be the result of deterministic manipulation. It begins by offering a Moorean-style proof of deliberation-compatibilism. It then turns to the leading argument for deliberation-incompatibilism, which is based on the presumed incompatibility of causal determinism and the 'openness' required for rational deliberation. The paper explains why this argument fails and develops a coherent account of how one can rationally deliberate and believe in causal determinism without inconsistency. The second half of the paper then takes up Cohen's proposed counterexample and his Four-Case Deliberation Argument (FCDA) against deliberation-compatibilism, which is meant to mirror Derk Pereboom's famous Four-Case Manipulation Argument. In response, the author defends a hard-line reply to FCDA but also argues that the notion of 'sourcehood' relevant to rational deliberation differs from that involved in free will.

Keywords: determinism, reason, rational deliberation, manipulation, compatibilism, free will, sourcehood.

Causal determinism is the thesis that every event and action, including human action, is the inevitable result of proceeding events and actions in conjunction with the laws of nature. Besides the usual threats that causal determinism poses to free will and basic desert moral responsibility (see Caruso 2012, 2021; Pereboom 2001, 2014), some philosophers maintain that *belief* in the truth of causal determinism also poses a threat to rational deliberation (see, e.g., Ginet 1966; Haji 2012; Kant 1785/1981; Taylor 1966). Those who find this threat illusory are deliberation-compatibilists:[1]

[1] The opening of this paper, including the definitions of deliberation-compatibilism and -incompatibilism, mirrors in form and content of the opening of Yishai Cohen's (2018) paper.

> **Deliberation-Compatibilism:** *S*'s deliberating and being rational is compatible with *S*'s believing that their actions are causally determined by antecedent conditions beyond their control.

Their opponents are deliberation incompatibilists:

> **Deliberation-Incompatibilism:** *S*'s deliberating and being rational is incompatible with *S*'s believing that their actions are causally determined by antecedent conditions beyond their control.

This paper aims to defend deliberation-compatibilism against several objections, including a recent counterexample by Yishai Cohen (2018) that involves a deliberator who believes that whichever action she performs will be the result of deterministic manipulation.

In Section I, I begin by offering a Moorean-style proof of deliberation-compatibilism and consider one possible reply. I then turn, in Section II, to the leading argument for deliberation-incompatibilism, which is based on the presumed incompatibility of causal determinism and the 'openness' required for rational deliberation. I explain why this argument fails and develops a coherent account of how one can rationally deliberate and believe in causal determinism without inconsistency. The account, as we'll see, draws heavily on previous work by Derk Pereboom (2014), Dana Nelkin (2004, 2011), and others, and offers an epistemic openness condition and an epistemic condition on the efficacy of deliberation. I then turn, in Section III, to Cohen's counterexample and spell out in detail his so-called *Four-Case Deliberation Argument* (FCDA) against deliberation-compatibilism, which is meant to mirror Pereboom's (2001, 2014) famous *Four-Case Manipulation Argument* (FCMA). In Section IV, I respond to Cohen by offering a 'hard-line' reply to his FCDA and in the process explore some deeper issues related to the notions of 'sourcehood' and 'control'.

I. MOOREAN-STYLE PROOF OF DELIBERATION-COMPATIBILISM

Deliberation is an essential component of action-guidance. It includes the ability to form and revise a conception of how we each wish to live, to conform behaviour to various goals and ends, and to deliberate among alternative means to achievement of those ends. It is a process or activity in which one figures out what to do. As Richard Taylor (1966: 168) notes, deliberation has 'as its aim or goal a decision to act', as opposed to the goal of merely forming a *belief* about which action one will perform. And as Yishai Cohen writes, 'Unlike the epistemic activity of inferring or predicting what will occur, deliberation is an activity or process that is intended to play an explanatory role with respect to what one ends up doing' (2018: 86). For the purposes of

this paper, I will adopt the following definition of deliberation offered by Derk Pereboom:

> **(D)** *S* deliberates just in case *S* is engaged in an active mental process whose aim is to figure out what to do from among a number of distinct, i.e. mutually incompatible, alternatives, a process understood as one that can (but need not) include the weighing and evaluating of reasons for the options for what to do. (2014: 110)

Rational deliberation, as distinct from deliberation simpliciter, requires that in addition to the above, the beliefs salient to an agent's deliberation be consistent. That is, in order to rationally deliberate about whether to do A1 or A2, where A1 and A2 are distinct actions, an agent must not have any inconsistent beliefs that are salient to her deliberation about whether to do A1 or A2 (Cohen 2018: 86). This means that I cannot rationally deliberate about whether I should walk home from campus today or use my powers of flight and fly, while also believing that I'm human and humans are incapable of flight.

The question under consideration in this paper is whether rational deliberation is compatible with belief in causal determinism. I maintain that it is. And in support of that claim I offer the following simple argument:

(1) If *S* can rationally deliberate among distinct actions A1. . . An *and* believe that their actions are causally determined by antecedent conditions beyond their control, then deliberation-compatibilism is correct—i.e., *S*'s deliberating and being rational is compatible with *S*'s believing that their actions are causally determined by antecedent conditions beyond their control.
(2) *S* can rationally deliberate among distinct actions A1. . . An *and* believe that their actions are causally determined by causal conditions beyond their control. [In fact, I did both this morning!]
(3) Therefore, deliberation-compatibilism is correct—i.e., *S*'s deliberating and being rational is compatible with *S*'s believing that their actions are causally determined by antecedent conditions beyond their control.

As proof of premise (2), the crucial premise, I offer the following pair of *Moorean facts*—facts I take to be more certain than any philosophical arguments to the contrary:

(a) I rationally deliberated this morning about what to wear. [I weighed multiple options, considered the weather, what I would be doing, who I might see, what looked best, and ultimately decided on the outfit I'm currently wearing.]
(b) I believed then and believe now that my actions are causally determined by antecedent conditions beyond my control.

I maintain that (a) and (b) together with the argument from modus ponens above provide a *Moorean-style proof of deliberation-compatibilism*.

Of course, like G.E. Moore's own 'proof' of the external world, where some zig others zag. As philosophers like to say, one person's *modus ponens* is another's *modus tollens*—and that's literally the case here. While I contend that premise (2) is obvious from the fact that I did both this morning—i.e., I engaged in an active mental process aimed at figuring out what to wear while also believing in causal determinism—deliberation-incompatibilists will deny premise (2) and replace my modus ponens with a modus tollens. Of course, they will provide arguments for why we should reject (2), but the question at hand is whether the arguments they provide are strong enough to overcome my two Moorean facts. For instance, they could argue that I'm simply mistaken about my belief in causal determinism. That is, while I may *believe that I believe* that my actions are causally determined by antecedent conditions beyond my control, I don't really *believe* what I *believe that I believe*. Such a move, however, is bound to fail for (at least) two main reasons. First, to simply assume I must be mistaken about (b) because it's obviously impossible to believe in determinism and rationally deliberation at the same time, would be to beg the question against the deliberation-compatibilist and assume the very thing that is under dispute. Secondly, while it may be possible for one to *believe that P* without also *believing that one believes that P*, it is not at all clear that the opposite is the case. That is, to *believe that one believes that P* would appear sufficient for one to *believe that P*. For these and other reasons, I'll assume this strategy is not the one deliberation-incompatibilists should take.

The more promising approach, and the one taken by most deliberation-incompatibilists, is to argue that while I may have deliberated about what to wear this morning, I did not *rationally* deliberate. That's because rational deliberation requires that I have no inconsistent beliefs salient to my deliberation, yet belief in determinism *is* inconsistent with a necessary condition for rational deliberation: the belief in the *openness* of options. In the following section, I'll examine this argument for deliberation-incompatibilism and explain why it too fails. I'll argue that deliberation only requires epistemic openness, not metaphysical openness, and epistemic openness does not conflict with an agent's believing that their actions are causally determined by antecedent conditions beyond their control. I'll further argue that deliberation-compatibilists should embrace an epistemic condition on the efficacy of deliberation in additional to an epistemic openness condition. I'll spell out my preferred formulations of each of these conditions and argue that they provide a coherent account of how one can rationally deliberate and believe in causal determinism without inconsistency.

II. RATIONAL DELIBERATION AND DETERMINISM

One of the main concerns' deliberation-incompatibilists have with determinism is that is appears to rule out the kind of *openness* of options required for rational deliberation. When we deliberate, we typically believe that we have

more than one distinct option available to us for which action to perform, each of which is available to us in the sense that we 'can' or 'could' perform each of these actions. It is often argued that belief in such openness is required for deliberation, or at least for rational deliberation. For example, Peter van Inwagen writes, 'if someone deliberates about whether to do A or to do B, it follows that his behavior manifests a belief that it is *possible* for him to do A—that he *can* do A, that he has it within his power to do A—and a belief that it is possible for him to do B' (Ginet 1966; cf. Kant 1785/1981: AK IV 448; Stapleton 2010; Taylor 1966: ch.12; van Inwagen 1983: 155). Some philosophers maintain that belief in this kind of openness conflicts with the truth of determinism in the sense that, in any deliberative situation, the truth of determinism would rule out the availability to us of all but one distinct option for what to do, and thus would rule out the openness about what to do. Accordingly, this line of reasoning supports a kind of deliberation-incompatibilism (see, e.g., Ginet 1966; Taylor 1966), which maintains that S's deliberating and being rational is incompatible with S's believing that their actions are causally determined.

But does determinism conflict with the kind of openness required for rational deliberation? Most deliberation-compatibilists acknowledge that deliberation requires a kind of openness, but rather than interpret it metaphysically they provide an epistemic interpretation of 'can' or 'could'. As Pereboom explains:

> It does seem plausible that when we deliberate about what to do, we typically presuppose that we have more than one distinct option for which action to perform, each of which is available to us in the sense that we can or could perform each of these actions. But the sense of 'can' or 'could' featured in such beliefs might not always or even typically be metaphysical. It might well be that in some such cases, it is epistemic, and in many others it is indeterminate between a metaphysical and epistemic sense. On certain epistemic interpretations, such beliefs would not conflict with a belief in determinism. When I am deliberating whether to do A, supposing I correctly believe determinism is true, I would not know whether I will in fact do A since I lack the knowledge of the antecedent conditions and laws that would be required to make the prediction based on these factors, not to mention the time and wherewithal. So even if I believe that it is causally determined that I will not do A, I might without inconsistency believe that it is in a sense epistemically possible that I do A, and that I could do A in this epistemic sense. (2014: 107)

Epistemic accounts of this kind have been developed by a number of deliberation-compatibilists, including Dennett (1984), Kapitan (1986), Pettit (1989), Nelkin (2004, 2011), and Pereboom (2014). The account I prefer maintains that the beliefs about the possibility of acting salient for deliberation are in some key respects epistemic but that there are *two* key compatibilist epistemic states. One of these specifies an epistemic notion of openness for what to do, and the other is an epistemic condition on the efficacy of deliberation (see, e.g., Kapitan 1986; Pereboom 2014). In what follows, I'll focus on Pereboom's formulations of these conditions since they plausibly deliver a coherent way of making sense of the relevant epistemic notions of openness and deliberative ef-

ficacy, while at the same time avoiding some of the more well-known counterexamples that have plagued other extant accounts (see, e.g., Pereboom 2014: ch.5).

The *epistemic openness* condition I endorse can be articulated as follows:

> **(EO)** In order to deliberate rationally among distinct actions A1. . . An, for each Ai, S cannot be certain of the proposition that she will do Ai, nor of the proposition that she will not do Ai; and either (a) the proposition that she will do Ai is consistent with every proposition that, in the present context, is settled for her, or (b) if it inconsistent with some such proposition, she cannot believe that it is. (Pereboom 2014: 113)[2]

This condition maintains that in order for an agent to deliberate rationally among distinct action A1. . . An, for each Ai, the agent cannot be certain of the proposition that they will do some action Ai, nor of the proposition that they will not do Ai. Furthermore, the proposition that they will do Ai must be consistent with every proposition that, in the present context, is settled for them—where, 'A proposition is settled for an agent just in case she believes it and disregards any uncertainty she has that it is true, e.g., for the purpose of deliberation' (2014: 133). Clause (b) is required because although there may be certain cases in which I can rationally deliberate about whether to do Ai even if in fact my doing Ai is inconsistent with a proposition I regard as settled in that context, 'it is crucial that I then not believe that it is inconsistent'—since, 'if I did believe this, it's intuitive that I couldn't rationally deliberate about whether to do A' (2014: 114).

It is exactly this epistemic sense of 'can' or 'could' that was implicit in my Moorean-style proof, since when I deliberated this morning about what to wear, although I may have been causally determined to decide as I did, I was neither certain that I would pick shirt* (*the shirt I'm currently wearing), nor certain that I would not. Furthermore, the proposition that I would choose shirt* was consistent with every other proposition that was settled for me in the context of my deliberation. (EO) therefore provides a plausible understanding of the kind of epistemic openness required for rational deliberation, and in no way conflicts with the belief that one's actions are causally determined by antecedent conditions beyond their control.

On its own, however, the epistemic openness condition does not provide a successful compatibilist account of rational deliberation. Belief in the efficacy of deliberation is required in addition. To see why, consider the following example provided by van Inwagen:

> [I]magine that [an agent] is in a room with two doors and that he believes one of the doors to be unlocked and other door to be locked and impassable, though he has no idea which is which: let him then attempt to imagine himself deliberating about which door to leave by. (1983: 154)

[2] In *Free Will, Agency, and Meaning in Life* (2014), Pereboom labels this principle (S). I am here relabeling it (EO)—for the *Epistemic Openness* condition—for sake of consistency.

About this example, Dana Nelkin remarks: 'While it seems that I can deliberate about which door to decide *to try* to open and even which door handle to decide to jiggle, if I know one of them to be locked and impassable, it also seems that I cannot deliberate about which *door to open*—or even which door to *decide* to open' (2011b: 130). Van Inwagen's example poses a problem for deliberation-compatibilists since it satisfies (EO) but is also plausibly a case where rational deliberation about which door to open is ruled out. What's more:

> . . . if an agent believed determinism and its consequences, then in any deliberative situation she would believe that all but one option for what to do was closed off; 'locked and impassible,' so to speak (although she would ordinarily not have a belief about which one was not closed off). If in the example one cannot deliberate about which door to open, and one believed determinism and its consequences, then it seems that one would never be able to deliberate about what to do. A compatibilist account would need to explain why rational deliberation is not possible in the two-door case, but nonetheless possible for the determinist. (Pereboom 2014: 116)

To solve this problem, several deliberation-compatibilists have suggested that rational deliberation also requires a belief in the efficacy of deliberation (see Kapitan 1986: 247; Nelkin 2004b; Pereboom 2014). That is, rational deliberators must believe that for each of the options for action under consideration, deliberation about it would, under normal conditions, be efficacious in producing the choice for that action and the action itself (Pereboom 2014: 117). In van Inwagen's example, then, we can say that it's not the absence of a belief in openness that precludes deliberation about which door to open. Rather, what precludes such deliberation is that given the agent's belief that one of the two doors is locked, if he is rational he will believe that his deliberation would not ultimately be efficacious for him opening one of the doors (Pereboom 2014: 117). This is not the case, however, in the normal case of determinism. That is, unlike the two-door case, when a determinist is deliberating under ordinary doxastic circumstances, he can, upon proper reflection, form the true belief that his deliberation makes a difference with respect to which action he performs. So there is an explanation for why the agent cannot rationally deliberate in the two-door case that does not apply to ordinary doxastic scenarios in which a determinist deliberates (Cohen 2018: 91).

We therefore need to add to (EO) a second *deliberative-efficacy condition*. Nelkin (2011: 142) and Kapitan (1996: 436) each offer formulations of their own, but I will once again focus on Pereboom's formulation:

> **(DE)** In order to rationally deliberate about whether to do A1 or A2, where A1 and A2 are distinct actions, an agent must believe that if as a result of her deliberating about whether to do A1 or A2 she were to judge that it would be best to do A1, then, under normal conditions, she would also, on the basis of this deliberation, do A1; and similarly for A2. (2014: 118-9)

The important thing to note is that while (DE) is not met by the agent in the two-door situation, it is satisfied by someone in an ordinary deliberative situation in which they believe that determinism is true and that they therefore have only one possibility for decision and action—but they do not know which. Hence, (DE) avoids van Inwagen's counterexample while making sense of the belief in deliberative efficacy under ordinary doxastic circumstances in which a determinist deliberates. If an agent believes that because determinism is true they cannot either do A1 or A2 on the basis of deliberation, but they do not know which, they can still meet condition (DE): for they might still rationally believe that if they were to judge doing A1 best, they would do A on the basis of deliberation, and similarly for A2.

Returning to my Moorean-style proof of deliberation-compatibilism, we can now say that (EO) and (DE), together with other uncontroversial conditions necessary for rational deliberation, provide a plausible and coherent account of how I can deliberate about what to wear *and* believe, without inconsistency, that my actions are causally determined by antecedent conditions beyond my control. Rational deliberation only requires epistemic openness and an epistemic condition on the efficacy of deliberation, neither of which conflict with believing in causal determinism. So, when I engaged in an active mental process aimed at figuring out what to wear this morning, my deliberation in no way conflicted with my belief in causal determinism, since I satisfied the epistemic openness condition (EO) and the deliberative-efficacy condition (DE).

But perhaps the deliberation-compatibilist is not out of the woods yet.

III. COHEN'S COUNTEREXAMPLE

Let me now turn to a recent objection by Yishai Cohen (2018) against the kind of account just sketched. Cohen offers, what he considers to be, a counterexample to 'all recent pro-*DC* [deliberative-compatibilist] views' (2018: 92). The putative counterexample involves an agent who satisfies all of the requirements for rational deliberation according to deliberation-compatibilists, including (EO) and (DE), and yet the agent apparently cannot rationally deliberate about what to do in light of her belief concerning her impending deterministic manipulation. A simplified version of Cohen's counterexample can be summarized as follows (Cohen 2018: 92–3): [Note: While Cohen's original presentation includes several premises addressing other leading formulations of the deliberative-efficacy condition by Kapitan (1996: 436), Clarke (1992: 103), Dennett (1984: 115), and Nelkin (2011: 142), I focus here only its treatment of Pereboom's formulation. I will grant that if the counterexample succeeds against (EO) and (DE), it succeeds *tout court* as an argument against 'all recent pro-DC views'.]

Case 1 (I)-(IV) are true:

(I) Betty believes the following. Betty is offered a choice to press one of the two buttons in front of her. If she presses the left button (henceforth 'LEFT'), Betty will receive $1000. If she presses the right button (henceforth 'RIGHT'), then Oxfam will instead receive $1000. Betty cannot alter who received the money once the first button is pressed. Moreover, if Betty presses both buttons simultaneously or presses no buttons at all, then no one received the money.

Betty is a U.S. citizen who is financially better off than most people in the world but is nevertheless burdened with financial debt. She has a strong desire to press LEFT in order to pay off some of her debt. However, Betty believes that donating the money to Oxfam will benefit people who are far worse off then her (Betty knows that she frequently but not exclusively undergoes rationally egoistic tendencies).

(II) Betty believes neither that she will press LEFT, nor that she will press RIGHT. Moreover, all of Betty's beliefs are consistent with the proposition that she will press LEFT, and the proposition that she will press RIGHT. *So Betty satisfies the condition in EO.*

(III) Betty believes that if as a result of her deliberation about whether to press LEFT or RIGHT she were to judge that it would be best to press LEFT, then, under moral conditions, she would also, on the basis of this deliberation press LEFT; and similarly for pressing RIGHT. *So Betty satisfied the condition in DE.*

(IV) Betty believes the following. A team of neuroscientists has the ability to manipulate her neural states at any time by radio-like technology. Prior to Betty's deliberation, the neuroscientists have decided arbitrarily (on the basis of a coin toss) to causally affect Betty's imminent decision (cf. Pereboom 2014: 76–7). As a result, the neuroscientists will manipulate Betty to press (and decide to press) one of the buttons by exerting either an egoism-enhancing or egoism-diminishing monetary influence upon Betty. If they exert a monetary egoism-enhancing influence, then Betty will press LEFT. If they exert a monetary egoism-diminishing influence, then Betty will press RIGHT.

While Betty does not know which kind of influence she will undergo, she believes that the neuroscientists only have the capability of *either* enhancing *or* diminishing Betty's egoistic tendencies. In other words, if the neuroscientists are capable of diminishing Betty's egoistic tendencies, then they do not have the capability to enhance such tendencies (and vice versa).

Finally, the neuroscientists will manipulate Betty's *decision* (which results from her deliberation) to press one of the buttons. The neuroscientists do not in any way alter Betty's ultimate *judgement* concerning what she has most

> reason to do, all things considered. So Betty's decision will be manipulated by slightly altering Betty's egoistic tendencies while Betty deliberates in order for her deliberation to generate a different 'output' than it might otherwise generate in the absence of such a manipulation.

Cohen further stipulates:

> Notice that (according to Betty's beliefs) the neuroscientists will in fact intervene, even if Betty would perform the same action in the absence of such manipulation. If I were to stipulate instead that the neuroscientists intervene only if, in the absence of manipulation, Betty would not have pressed the neuroscientists' pre-selected button, then *Case 1* would strike a resemblance with Frankfurt-style cases, which in turn would raise numerous vexing issues that are beyond the scope of this paper. For this reason, I maintain that the intervention by the neuroscientists does not depend upon what Betty would do in the absence of manipulation. Moreover, we may also stipulate that Betty has no belief about what she would in fact do in the absence of this manipulation since, according to (I), she believes that she frequently but not exclusively undergoes rationally egoistic tendencies. (2018: 93-4)

With this understanding of *Case 1*, let us inspect whether it is a genuine counterexample to the account of deliberation-compatibilism sketched in the previous section.

Cohen provides the following argument for why the counterexample succeeds:[3]

(1*) In *Case 1*, Betty satisfies the conditions in EO and DE with respect to rationally deliberating about which button to press.

(2*) In *Case 1*, Betty satisfies the *no inconsistent beliefs condition* (which Cohen calls the (NIB) thesis) with respect to rationally deliberating about which button to press. [NIB maintains that in order to rationally deliberate about whether to do A1 or A2, where A1 and A2 are distinct actions, an agent must not have any inconsistent beliefs that are salient for her deliberating about whether to do A1 or A2 (Cohen 2018: 86).]

(3*) In *Case 1*, Betty cannot rationally deliberate about which button to press.

(4*) If (1)-(3) are true, then *Case 1* is a counterexample to deliberation-compatibilism and the account defended in the previous section.

(5*) Therefore, *Case 1* is a counterexample to deliberation-compatibilism and the account defended in the previous section.

Since premises (1*) and (4*) are uncontroversial, especially since (1*) was stipulated as true as part of Cohen's *Case 1*, I'll simply grant these. The

[3] Again, I'm simplifying here by leaving out the various conditions offered by other leading deliberation-compatibilist accounts, which Cohen stipulates his counterexample also satisfies.

premises in need of defence are premises (2*) and (3*). And while I think there are important issues worth exploring with regard to premise (2*), since this is one potential place to attach the argument, I'll grant this premise as well so as to focus my attention on premise (3*).

Cohen argues for premise (3*) on the basis of the following principle:

***Causal Influence*:** Necessarily, if an agent *S* believes the following,

- Either agent *T* will φ or *T* will ψ (but *T* will not perform both actions),
- *S* cannot causally contribute to either *T*'s φ-ing or *T*'s ψ-ing.
- *T*'s φ-ing is (in conjunction with the laws of nature) causally sufficient for the occurrence of event *e*.
- *T*'s ψ-ing is (in conjunction with the laws of nature) causally sufficient for the occurrence of event *e*.

then *S* cannot rationally deliberate about whether to permit the occurrence of *e*.

In order to motivate this principle, Cohen provides the following example:

> Alex is viewing a live television broadcast of an eight-ball billiards match. The opening move is made, but no balls are pocketed. The next player will strike the cue ball towards the left or toward the right. In conjunction with the laws of nature, striking the cue ball towards the left is causally sufficient for the pocketing of the #14 striped ball, and striking the cue ball towards the right is causally sufficient for the pocketing of the #7 solid ball. Viewing this match from home, Alex cannot causally contribute to the player's next move. Moreover, Alex believes all of this. So Alex cannot rationally deliberate about whether to permit the pocketing of the #14 striped ball or the #7 solid ball. Generalizing from this case, it follows that *Causal Influence* is true. (2018: 94-5)

Given the truth of *Causal Influence*, Cohen argues that we can establish premise (3*) once we recall what Betty believes according to (IV). That is, either the neuroscientists will decide that Betty presses LEFT or the neuroscientists will decide that Betty presses RIGHT. She cannot causally contribute to the neuroscientists' nefarious activity (recall that their decision is based on a coin toss). The neuroscientists' decision that Betty pressed LEFT is (in conjunction with the laws of nature) causally sufficient for the occurrence of Betty's pressing LEFT (and similarly for RIGHT). Cohen therefore concludes: 'It thus follows from *Causal Influence* that Betty cannot rationally deliberate about whether to permit the occurrence of Betty's pressing LEFT or Betty's pressing RIGHT. So Betty cannot rationally deliberate about which button to press' (2018: 95).

After providing this defence of premise (3*), Cohen proceeds to offer an FCDA against deliberation-compatibilism similar to Pereboom's FCMA against the kind of compatibilism relevant to the free will debate—i.e., the compatibility of determinism and the kind of free will required for basic desert moral responsibility (see Pereboom 2001, 2014). According to ***Case 2***, premises (I)–(III) are all the same and again true, but instead of (IV), (V) is true:

(V) Betty believes the following. Long ago, a team of neuroscientists decided arbitrarily (on the basis of a coin toss) which button Betty is to press (and decide to press). As a result, these neuroscientists have programmed Betty at the beginning of her life in such a manner that she will press (and decide to press) one of the buttons, though Betty has no belief about which button the neuroscientists want her to press. (Cohen 2018: 98–9)

According to ***Case 3***, (I)–(III) are the same, but instead of (IV), (VI) is true:

(VI) Betty believes the following. The training practices of Betty's community (which were completed before she developed the ability to prevent or alter these practices) causally determined the nature of her deliberative reasoning processes such that, in conjunction with certain background conditions, Betty is causally determined to press (and decide to press) one of the buttons. Though, Betty has no belief about which button she will in fact press. (2018: 99)

Lastly, according to ***Case 4,*** (I)–(III) are the same, but instead of (IV), (VII) is true:

(VII) Betty believes the following. Everything that happens in the universe is causally determined by its past states together with the laws of nature. Betty is an ordinary human being raised in normal circumstances. As a result, Betty's deliberative reasoning processes, in conjunction with certain background conditions, will causally determine Betty to press (and decide to press) one of the buttons. Though, Betty has no believe about which button she will in fact press (Cohen 2018: 99).

Cohen maintains that whether Betty believes that the process of manipulation begins a few seconds prior to her decision (in *Case 1*) or at the beginning of her life (in *Case 2*) does not make a difference with respect to Betty's ability to rationally deliberate about which button to press. The same is true with respect of Betty's beliefs in *Case 2* and *3*; and similarly for *Cases 3* and *4*. This leads Cohen to conclude:

> So in light of the fact that Betty cannot rationally deliberate in *Case 1*, Betty cannot rationally deliberate in any of these four cases. The best explanation for Betty's inability to rationally deliberate in all four cases is that Betty believes that whichever action she performs (and decides to perform) will be causally determined by factors beyond her control. (2018: 99)

Hence, according to Cohen, rational deliberation requires an agent to lack the belief that her action will be causally determined by factors beyond her control. And as a result, rational deliberation is incompatible with belief in causal determinism.

But Cohen also suggests that there is a 'more fundamental explanation as to why [deliberation-compatibilism] is false', namely: 'Perhaps one must believe that one will be the source of one's action, such that being the source of one's action is incompatible with determinism (Kant 1785/1981, 448; Taylor 1964, 76; Castañeda 1975, 134–135)' (Cogen 2018: 99). But instead of proposing that an agent must believe some proportion p, Cohen claims that the deliberation-incompatibilist can resort to the weaker claim that an agent must *lack* the belief that *not*-p. He thus offers the following requirement on rational deliberation:

> **Source:** In order to rationally deliberate about whether to do A1 or A2, where A1 and A2 are distinct actions, an agent S must not believe that it is not the case that *S* will be the source of whichever action *S* performs, such that being the source of one's action is incompatible with the action being causally determined by factors beyond one's control. (2018: 99)

He further suggests that the notion of 'sourcehood' relevant to rational deliberation is the same notion relevant to moral responsibility, such that Pereboom's FCMA against basic desert moral responsibility is sound if and only if Cohen's FCDA against deliberation-compatibilism is sound (2018: 101).

IV. WHY COHEN'S MANIPULATION ARGUMENT IS NOT A COUNTEREXAMPLE

Cohen's FCDA is an interesting one. It also poses a unique challenge for someone like me, who accepts deliberation-compatibilism but who also believes that manipulation arguments, like Pereboom's FCMA, succeed in refuting compatibilism in the free will debate (see Caruso 2014, 2021; Dennett and Caruso 2021; Pereboom and Caruso 2018). In this section, I'll argue that Cohen is mistaken that FCDA and FCMA 'stand or fall together' (2018: 101), and that a proper understanding of the notions of 'sourcehood' and 'control' relevant to rational deliberation will allow us to distinguish them from the kind of control in action, i.e., free will, required for basic desert moral responsibility. Furthermore, I'll offer a 'hard-line' reply to Cohen's FCDA, challenging (3*) and arguing that Betty can, in fact, rationally deliberate about which button to press. My response will be based on an important disanalogy between Betty and Alex.

To begin, there are essentially two ways to respond to Cohen's FCDA argument, which divide into so-called *hard-line* and *soft-line* replies. The hard-line reply claims that in *Case 1*, Betty *can* in fact rationally deliberate about which button to press, hence premise (3*) of Cohen's argument is false (Haas 2013; cf. McKenna 2008, 2014). The hard-line reply rejects the central intuition that the kinds of manipulation involved in *Cases 1, 2*, and *3* undermine rational deliberation. It recommends that instead of beginning with *Case 1* and working toward the case of natural determinism, we work our intuitions in the opposite

direction. That is, the hard-line reply begins with *Case 4*, the case of natural determinism, and concludes that in such circumstances Betty can rationally deliberate about which button to press. It then argues that, since there are no relevant differences between *Cases 4* and *3*, *3* and *2*, and *2* and *1*, capable of justifying a different outcome, we should conclude that Betty can rationally deliberate about which button to press in *Case 1*, despite the truth of (IV).

The soft-line reply, on the other hand, seeks to identify a relevant difference between two adjacent cases, such that Betty can rationally deliberate in one of these cases but not the other. For example, one could argue that the reason Betty cannot rationally deliberate in *Cases 1–3* but can in *Case 4* is that, it is only in the latter case that Betty does not believe that the causal determination of her choice includes in some manner the intentional actions of others (cf. Lycan 1997: 115–9). Or, perhaps, one could argue that since Betty believes that she is being manipulated in a particularly invasive manner only in *Case 1* (cf. Demetriou 2010; Fischer and Tognazzini 2011: 18–25) or only in *Cases 1–2* (cf. Mele 2006: 141–4), Betty cannot rationally deliberate in these cases. Soft-line replies therefore grant that Cohen is correct that in *Case 1* Betty cannot deliberate about which button to press, but they deny that this conclusion carries over to the case of natural determinism (*Case 4*). According to soft-liners, while Cohen's *Case 1* may establish that rational deliberation is not compatible with the belief that our neural states have been manipulated by a team of neurosciences, it does not establish that rational deliberation is incompatible with a general belief in natural determinism.

I'll explain below why I think the hard-line reply is not only the correct response, but why it succeeds against Cohen's FCDA but not as a reply to Pereboom's FCDA.

Against the hard-line response, Cohen writes: 'The main concern with this response is that one must deny the *Causal Influence* principle since this principle entails premise [3*]' (2018: 100). Cohen's *Causal Influence* principle, recall, was motivated by the eight-ball billiards match example. According to Cohen, the principle renders the correct verdict that Alex cannot rationally deliberate about whether to permit the pocketing of the #14 striped ball or the #7 solid ball. The problem, however, is that there is an important disanology between Alex (in the eight-ball billiards example) and Betty (in the FCDA). Namely, *Betty is in a position to contribute causally to what happens next, while Alex is not.* Betty, recall, not only satisfies the conditions of epistemic openness and deliberative-efficacy, she also satisfies other uncontroversial conditions on agency. For instance, Betty's *decision* to press one of the buttons, regardless of whether it is manipulated by a team of neuroscientists or causally determined by natural factors outside her control, causally contributes to her pressing the button. For instance, if her egoistic tendencies are momentarily enhanced, due either to natural deterministic factors or manipulation, she will decide and then press LEFT. If, on the other hand, her egoistic tendencies are momentarily

diminished, she will decide and then press RIGHT. Either way, her agential structures will play an important causal role in which button she presses. This is not the case with Alex. In fact, Alex exercises no causal influence over which billiard ball is pocketed nor does he believe that he does. In Alex's case, the relevant causation does not pass through his agential structure.

Cohen appears to be aware of this type of reply, since he writes: 'A hard-liner who seeks such an alternative explanation could modify the last part of *Casual Influence* in the following manner: "... then *S* cannot rationally deliberate about whether to permit the occurrence of *e*, *unless e concerns a decision by S*"' (2018: 100). Cohen acknowledges that: 'This modified principle can account for Alex's inability to rationally deliberate, and can accommodate the position that Betty can rationally deliberate in *Case 1*' (2018: 100). Since Cohen acknowledges that this modification would sink his entire argument by providing a straightforward explanation of why Alex is unable to rationally deliberate, but Betty is, he must reject it. So how does he go about arguing against it? All we get is the following, rather unsatisfying two sentences: 'While I find this modification to *Causal Influence* to be ad hoc, I don't expect deliberation compatibilists to share this intuition. Breaking this stalemate in the future will require further dialectical maneuvers' (2018: 100).

Cohen is correct that I fail to find such a modification ad hoc. I don't find it ad hoc at all to think that Betty's *decision* to press LEFT or RIGHT, even if determined, is relevant to Causal Influence. Since Betty is in a position to contribute causally to what happens next, while Alex is not, Betty (though not Alex) is able to exercise causal influence over what happens next. I therefore maintain that the reason Alex is unable to deliberate about which ball to pocket next is that he cannot contribute causally to the next move in the sense that the causal chain sufficient for either the pocketing of the #7 ball or the #14 ball never 'passes through' him in a way that would allow him to causally contribute to the outcome and hence exercise causal influence. Betty, on the other hand, does exercise causal influence over what button she presses in the sense that her pressing LEFT or RIGHT involves a decision and concerns a causal chain that passes through her agential structures. I therefore agree with Cohen that an agent cannot rationally deliberate about events and actions that they have no causal influence over, but I strongly disagree that Betty lacks causal influence in anything like the same way that Alex does.

I propose, then, that for an agent to exercise the kind of causal influence relevant to rational deliberation, the causal chain must pass through their agential structures. We can call this the *deliberative causal influence* (DCI) principle and define it more accurately as follows:

> **(DCI):** For an agent *S* to exercise the kind of causal influence relevant to rational deliberation, the causal chain must pass through *S*'s agential structures and the resulting

> action, whatever it ends up being, be the result of *S*'s decision—where 'be the result of' is understood causally.

(DCI) allows us to identify an intuitively relevant difference between Betty and Alex—namely Betty satisfies DCI, while Alex does not. This difference, I contend, explains why Alex is unable to rationally deliberate in the billiards example—e.g., he believes he lacks deliberative causal influence over the next move. Betty, on the other hand, does not believe she lacks the relevant kind of deliberative causal influence, since (IV) does not undermine such causal influence, hence she can rationally deliberate. If this is correct, then Cohen has failed to provide a convincing argument for (3*), the crucial premise. My hard-line reply therefore maintains that Betty can rationally deliberate about which button to press in all four cases of Cohen's FCDA, as long as she satisfies (EO) and (DE), and also not believe that she lacks the kind of deliberative causal influence specified by (DCI).

Cohen may reply that while Betty satisfies (DCI) and Alex does not, what makes the two cases analogous is that Betty cannot causally contribute to the *neuroscientists'* nefarious activities, which are causally sufficient for the occurrence of Betty pressing LEFT or RIGHT, and she believes this. That is, Cohen could argue that the kind of causal influence relevant to rational deliberation requires that Betty be able to causally influence those distal causes (or some relevant subset of them) that are sufficient for her pressing LEFT or RIGHT. But that is exactly the claim I have just challenged. The mistake, I contend, is that such a demand conflates the kind of control necessary for free will and basic desert moral responsibility, with the kind of control required for rational deliberation. Rational deliberation, I contend, requires only the minimal kind of control specified in (DCI), and perhaps some other non-controversial conditions. The control in action required for free will, on the other hand, arguably requires Cohen's more robust notion of control. Once we separate these two notions, however, and see that rational deliberation only requires belief in the weaker notion of control, the problem goes away.

Let me conclude with some final thoughts on the deeper issues of *sourcehood* relevant to the free will debate. While I contend that Betty is able to rationally deliberate about which button to press in *Cases (1)–(4)*, without inconsistency, I also maintain that Pereboom's FCMA succeeds as an argument against compatibilism. I therefore disagree with Cohen that FCDA and FCMA 'stand or fall together' (2018: 101). Our disagreement, I think, comes down to the following question: Are the notions of control and sourcehood relevant to rational deliberation the same notions relevant to free will and moral responsibility? Cohen contends they are (2018: 101–3), while I contend they are not.

Consider, again, Cohen's *Source* principle, which maintains that: 'In order to rationally deliberate about whether to do A1 or A2, where A1 and A2

are distinct actions, an agent *S* must not believe that it is not the case that *S* will be the source of whichever action *S* performs'. He then defines sourcehood in incompatibilist terms, writing: 'being the source of one's actions is incompatible with the action being causally determined by factors beyond one's control' (2018: 99). But why define sourcehood in this way? By doing so, Cohen is lead to conclude that rational deliberation is incompatible with belief in causal determinism—since an agent must believe they are the source of whichever action they perform (or, at least, not believe they are not the source) in order to rationally deliberate, and such a belief (according to Cohen) is incompatible with causal determinism. He then adds, 'the same notion of sourcehood is relevant to both moral responsibility and rational deliberation' (2018: 101). That is: 'An agent must be the source of her action in order to be morally responsible for that action. Moreover, according to *Source*, in order to rationally deliberate, an agent must refrain from believing that she is not the source of whichever action she will in fact perform' (2018: 101).

I disagree with Cohen that the same notions of sourcehood are relevant to rational deliberation and moral responsibility. While I cannot provide a full accounting of the difference between these notions here, I will offer the following brief proposal which I believe provides a prima facie case for distinguishing the two notions. I propose that the kind of sourcehood relevant to rational deliberation is the kind captured by the principle of *deliberative causal influence* (DCI), which requires that an agent *S* causally contribute to whatever action *S* performs, such that the causal chain pass through *S*'s agential structures and the resulting action, whatever it ends up being, be the result of *S*'s decision. This provides an intuitively plausible account of the kind of sourcehood relevant to rational deliberation without assuming anything about the kind of control in action, i.e., free will, required for basic desert moral responsibility. Without begging the question, I see no reason why we should demand more than this *weak notion* of sourcehood when it comes to rational deliberation. Furthermore, this notion of sourcehood in no way conflicts with causal determinism, so there is no inconsistency in Betty not believing that she lacks such sourcehood (to stick with Cohen's way of formulating things) and Betty believing that her actions are causally determined by conditions beyond her control.

On the other hand, I propose that the kind of sourcehood and control required for free will is the kind needed for *basic desert* morally responsible (Caruso 2021; Dennett and Caruso 2021; Pereboom 2001, 2014). As Pereboom explains:

> For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary. The desert at issue here is basic in the sense that the agent would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialists or contractualist considerations. (2014: 2)

Understood this way, free will is a kind of power or ability an agent must possess in order to justify certain kinds of desert-based judgements, attitudes, or treatments—such as resentment, indignation, moral anger, and retributive punishment—in response to decisions or actions that the agent performed or failed to perform. These reactions would be justified on purely backward-looking grounds, that's what makes them *basic*, and would not appeal to consequentialist or forward-looking considerations, such as future protection, future reconciliation, or future moral formation (see Caruso and Morris 2017; Dennett and Caruso 2021; Pereboom 2001, 2014).

But once we define free will as the control in action required for basic desert moral responsible, it becomes at least prima facie plausible to think that free will requires something more than just the deliberative causal influence of agents. What this 'something more' is, is, of course, a matter of great dispute. But consider the following argument for free will incompatibilism from Peter van Inwagen:

> If determinism is true, then there is some state of the world in the distant past P that is connected by the laws of nature to any action A that one performs in the present. But since no one is responsible for the state of the world P in the distant past, and no one is responsible for the laws of nature that lead from P to A, it follows that no one is responsible for any action A that is performed in the present. (1983: 182-3)

This argument nicely captures one of the main incompatibilist intuitions about determinism and free will (see also Pereboom 2001: 34). The problem is that if determinism is true, then there are conditions for which no one is, or ever has been, even partly responsible (in the basic desert sense relevant to free will), and these conditions determine the *actual sequence that brings about* the agent's action. For reasons such as these, I have elsewhere argued for the following incompatibilist intuition about determinism and free will: 'An action is free in the sense required for [basic desert] moral responsibility only if it is not produced by a deterministic process that traces back to causal factors beyond the agent's control' (Pereboom 2001: 34; see, e.g., Caruso 2021; Dennett and Caruso 2021).

We can see, then, that there are other notions of sourcehood, *stronger notions*, which can be taken as the relevant notion in the traditional free will debate. For instance, source incompatibilists about free will would argue that the relevant notion of sourcehood requires that an agent *S* not only causally contribute to whichever action *S* performs, understood as defined by DCI, but that *S* also be the source of their action in a sense incompatible with *S*'s being causally determined by antecedent conditions beyond their control. *Agent-causal* libertarians may go further and add that *S* must also poses the control in action required to *settle* which action they perform, where such settling is not only incompatible with causal determinism but also incompatible with simple event indeterminism. Note, though, that under the assumption of determinism, weak sourcehood is still possible, since it in no way conflicts with determinism, while the stronger notion of sourcehood, what we might call *ultimate sourcehood*,

would be ruled out. So even if we grant Cohen the assumption that causal determinism is incompatible with ultimate sourcehood, a view I actually share, it does not follow that Betty must give up the belief that she is the source of her actions in the weak sense—the sense relevant to rational deliberation.

This all brings me to my final point. While I have offered a hard-line reply to Cohen's FCDA, it does not follow that a similar hard-line reply will work against Pereboom's FCMA. If I'm correct that the notions of sourcehood and control relevant to rational deliberation are distinct from the notions relevant to free will and basic desert moral responsibility, then it's possible for an incompatibilist in the free will debate to argue that causal determinism undermines the control in action required for basic desert moral responsibility, while at the same time claiming that the control required for rational deliberation remains compatible with causal determinism. This, in fact, is my own view. And although I have not argued for Pereboom's FCMA here, nor have I argued against the hard-line replies to it, I have done so elsewhere (Caruso 2021). There is no inconsistent belief, however, in defending this combination of views since *actual* manipulation by a team of neuroscientists may in fact pose a threat to the control in action required for basic desert moral responsibility, while *belief* in such manipulation *not* threaten the weaker notion of control required for rational deliberation. For this reason, FCMA and FCDA need not stand or fall together.

REFERENCES

Caruso, G. D. (2012) *Free Will and Consciousness: A Deterministic Account of the Illusion of Free Will*. New York: Lexington Books.

——— (2014) 'Precis of Derk Pereboom's Free Will, Agency, and Meaning in Life', *Science, Religion, and Culture*, 1(3): 178–201.

——— (2021) *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*. New York: Cambridge University Press.

Caruso, G. D. and Morris, S. G. (2017) 'Compatibilism and Retributive Desert Moral Responsibility: On What Is of Central Philosophical and Practical Importance', *Erkenntnis*, 82: 837–55.

Castañeda, H.-N. (1975) *Thinking and Doing*. Dordrecht: D. Reidel Publishing.

Clarke, R. (1992) 'Deliberation and Beliefs About One's Abilities', *Pacific Philosophical Quarterly*, 73: 101–13.

Cohen, Y. (2018) 'Deliberating in the Presence of Manipulation', *Canadian Journal of Philosophy*, 48: 85–105.

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

Dennett, D. C. (1984) *Elbow Room*. Cambridge, MA: MIT Press.

Dennett, D. C. and Caruso, G. D. (2021) *Just Deserts: Debating Free Will*. New York: Polity Books.

Fischer, J. M. and Tognazzini, N. A. (2011) 'The Physiognomy of Responsibility', *Philosophy and Phenomenological Research*, 82: 381–417.

Ginet, C. (1966) 'Might we have no choice?' In Lehrer, K. (ed.) *Freedom and Determinism*, 87–104. New York: Random House.

Haas, D. (2013) 'In Defense of Hard-Line Replies to the Multiple-Case Manipulation Argument', *Philosophical Studies*, 163: 797–811.

Haji, I. (2012) *Reason's Debt to Freedom*. New York: Oxford University Press.

Kant, I. (1785/1981) *Grounding for the Metaphysics of Morals*,Ellington, T. J., ed. Indianapolis, IN: Hacket.
Kapitan, T. (1986) 'Deliberation and the Presumption of Open Alternatives', *Philosophical Quarterly*, 36: 230–51.
——— (1996) 'Modal Principles in the Metaphysics of Free Will', *Philosophical Perspectives*, 10: 419–45.
Lycan, W. G. (1997) *Consciousness*. Cambridge, MA: MIT Press.
McKenna, M. (2008) 'Say Good-Bye to the Direct Argument the Right Way', *Philosophical Review*, 117(3): 349–84.
——— (2014) 'Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin', *Philosophy and Phenomenological Research*, 89: 467–84.
Mele, A. (2006) *Free Will and Luck*. New York: Oxford University Press.
Nelkin, D. K. (2004) 'Deliberative Alternatives', *Philosophical Topics*, 32(1): 215–40.
Nelkin, D. K. (2011) *Making Sense of Freedom and Responsibility*. New York: Oxford University Press.
Pereboom, D. (2001) *Living Without Free Will*. New York: Cambridge University Press.
——— (2014) *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
Pereboom, D. and Caruso, G.D. (2018) 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life', *In Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, ed. G. D. Caruso and O. Flanagan, pp.193-222. New York: Oxford University Press.
Pettit, P. (1989) 'Determinism and Deliberation', *Analysis*, 49: 42–4.
Stapleton, S. (2010) *Hard Incompatibilist Challenges to Morality and Autonomy*. Dissertation. Department of Philosophy, Cornell University, Ithaca NY.
Taylor, R. (1964) 'Deliberation and Foreknowledge', *American Philosophical Quarterly*, 1(1): 73–80.
——— (1966) *Action and Purpose*. Englewood Cliffs, NJ: Prentice-Hall Inc.
Van Inwagen, P. (1983) *An Essay on Free Will*. New York: Oxford University Press.

SUNY Corning, USA