What the Consequence Argument is an Argument For

Justin A. Capes

The consequence argument is among the most influential arguments for incompatibilism, the thesis that free will and determinism are mutually exclusive. Recently, however, it has become increasingly clear that even if the argument is sound, it nevertheless fails to establish that particular incompatibilist conclusion. Be that as it may, I'll argue that, if sound, a version of the argument does establish another important incompatibilist thesis about free will, one that many incompatibilists seem to have had in mind all along. This other thesis, we'll see, isn't simply a restricted version of incompatibilism and is arguably the more central incompatibilist position.

There are several different versions of the consequence argument. Here I'll focus on what is probably the most widely discussed version, van Inwagen's Third Formal Argument.[1] What I have to say about it, however, should be applicable, *mutatis mutandis*, to most other versions of the argument as well. The Third Formal Argument employs the following abbreviations: '□' stands for broad logical necessity, '⊃' for the material conditional, 'P_o' is a proposition describing the intrinsic state of the universe in the distant past before human beings existed, 'L' is the conjunction of all the laws of nature, 'P' is any true sentence, and, finally, 'Np' abbreviates 'p, and no one has, or ever had, any choice about whether p.' Here are two inference rules alleged by proponents of the argument to govern the use of this 'N' operator:

      (α) $\Box p \vdash Np$.

---

[1] See van Inwagen 1983, pp. 93-104.

(β) N $(p \supset q)$, N$p \vdash$ Nq.[2]

And here now is the Third Formal Argument:

      (1) $\Box$ ((P_o & L) $\supset$ P)   formal consequence of determinism

      (2) $\Box$ ($P_o \supset$ (L $\supset$ P))   from 1 by exportation

      (3) N ($P_o \supset$ (L $\supset$ P))   from 2 by rule α

      (4) N P_o premise

      (5) N (L $\supset$ P)   from 3 and 4 by rule α

      (6) N L   premise

      (7) N P   from 5 and 6 by rule β

Does this argument, which seems to show that we can derive the conclusion that no one has a choice about anything from the assumption that determinism is true, support incompatibilism about free will and determinism? Arguably, it doesn't, and here's why.

      Incompatibilism, understood as the thesis that free will and determinism are mutually exclusive, is a necessary truth, if true at all. It's the thesis that, necessarily, no one has free will if determinism is true, that no deterministic worlds contain even a single agent who is possessed of free will. The Third Formal Argument, like every other consequence-style argument of which I'm aware, attempts to establish that thesis by assuming determinism is true and deriving from that assumption the conclusion that no one has free will. But as Warfield (2000, p. 169) points

---

[2] It's widely recognized that rule β is invalid as it stands. There are, however, ways of modifying the principle so that it may well turn out to be valid. Fortunately, these niceties needn't concern us here. See Huemer 2000, McKay and Johnson 1996, and Widerker 1987 for discussion of the issue.

out, to show that determinism strictly implies the non-existence of free will using this conditional proof strategy, "one is restricted, on pain of modal fallacy, from appealing to merely contingent truths in between" the assumption of determinism and the conclusion that no one has free will. The Third Formal Argument violates this formal restriction, as two of its steps (4 and 6) are contingent truths, if true at all. The argument therefore doesn't establish incompatibilism. At best, it supports the weaker claim that if determinism is true, then no one has free will.[3]

We can elaborate on the general problem just adumbrated by considering step 4 of the Third Formal Argument. It says, "P_o, and no one has, or ever had, a choice about whether P_o." A truth, to be sure, but not a necessary one, since P_o is only contingently true. Nor is this difficulty easily circumvented by augmenting the argument with a premise to the effect that "necessarily, if determinism is true, there is a true description of the remote past P_r such that NP_r," for this auxiliary premise is controversial, at best. It's debatable whether having a remote past (i.e., a past prior to the existence of human agents) is an essential feature of deterministic worlds.[4]

Similar remarks apply to step 6 of the Third Formal Argument, which says, "L, and no one has, or ever had, a choice about whether L." Another truth, perhaps, but arguably not a necessary one, since the laws of nature are widely thought to be contingent truths. And here too the worry can't be easily sidestepped by supplementing the argument with a premise to the effect that, "necessarily, no one has a choice about laws of nature, whatever they may be," for this auxiliary premise is likely false. To see this, consider what Cutter (2017) refers to as coy miracle workers, agents who have "the ability to act in such a way that the laws of nature are violated,"

---

[3] Bailey 2012, Campbell 2007, Cutter 2017, and Warfield 2000 advance versions of this general objection.

[4] See Campbell 2007 and Sartorio 2015 for objections to this auxiliary premise. See also Bailey 2012, who points out that Campbell's "no past" objection applies to most arguments for incompatibilism, including most other versions of the consequence argument, since these other arguments, no less than the Third Formal Argument, rely in one way or another on the assumption that there is a remote past over which we have no control.

but who never exercise that ability. Coy miracle workers seem possible. "After all," Cutter says, "there are all sorts of bizarre things out in the far reaches of modal space" (2017, p. 8).[5] But if they are possible, it isn't a necessary truth that no one has a choice about laws of nature.[6]

What are we to make of these criticisms of the consequence argument. One of the most common responses is to argue that even if the argument fails to establish incompatibilism, it can still be used to support restricted versions of that thesis according to which determinism is incompatible with free will in agents relevantly like us—agents for whom there is a distant past and who aren't coy miracle workers, for example. There is, however, a potential difficulty with this response. According to Cutter (2017, p. 279), the conclusion of the consequence argument merits the name 'incompatibilism' only if it "would not be endorsed by paradigmatic compatibilists." As I understand him, Cutter isn't just wrangling about words here. His point, I take it, is that the consequence argument will help advance the debate between compatibilists and incompatibilists only if it can be used to establish a thesis with which paradigmatic compatibilists (qua compatibilists) would disagree. But, as Cutter goes on to argue, it's unclear whether the restricted theses the consequence argument is said to support satisfy this constraint. Consider, for example, the restricted incompatibilist thesis that, necessarily, if determinism is true, then no one who isn't a coy miracle worker has free will. As Cutter (2017, p. 285) points out, local miracle compatibilists like David Lewis "could happily accept" this thesis, since, according to their brand of compatibilism, we might be free to do otherwise even if determinism

[5] Depending on how we should view laws of nature, we may not need to travel to "the far reaches of modal space" to find coy miracle workers. These beings might be sitting in our chairs and thinking our thoughts. That is, *we* might be coy miracle workers! This might be true, for example, on a Humean view of the laws. For a discussion of this point in relation to the consequence argument, see Beebee and Mele 2008. For a discussion of the point in relation to incompatibilist views more generally, see Rosen 2002, pp. 705-706.

[6] See Stone 1998 for a related objection to step 6.

is true, though, if we had done otherwise, an actual law of nature would have been violated.[7] That the consequence argument can be used to establish this thesis therefore doesn't help advance the debate between incompatibilists and at least one prominent group of compatibilists, and Cutter is skeptical that the consequence argument supports *any* restricted version of traditional incompatibilism that would not be accepted by paradigmatic compatibilists (p. 286).

Suppose Cutter is right about that. Suppose, more generally, that the aforementioned criticisms of the consequence argument are correct. What lessons should we draw from this? Should we conclude that the argument is irrelevant to the issue of free will? Should we conclude, as Cutter does, that the argument "does not support any thesis that could reasonably be called 'incompatibilism'" (2017, p. 278)? We should not, as I'll now argue. A bit more precisely, I'll argue that a version of the consequence argument supports a philosophically interesting conclusion about free will, one that could reasonably be called incompatibilism. This conclusion, which I'll call incompatibilism*, says that, necessarily, if all of an agent's behavior is necessitated by circumstances over which the agent has no control, then the agent lacks free will.

Incompatibilism* isn't an insignificant thesis. It entails, for example, that if, as many believe, the past and the laws of nature are indeed beyond our control (even if the laws aren't *necessarily* beyond our control), and if, as some believe, our actions are all necessitated by events in the distant past in conjunction with the laws of nature, then none of us has free will. It also entails that if, as some theologians believe, our actions are the inevitable upshot of God's will (something over which we presumably have no control), then none of us has free will.

Notice also that incompatibilism* isn't simply a restricted version of the traditional incompatibilist position. If it were, it would be entailed by that position, which it isn't (though

---

[7] See Lewis 1981.

it's difficult to see why someone would accept incompatibilism but not incompatibilism*).[8] Nor does incompatibilism* entail the incompatibility of free will and determinism (at least not in the absence of controversial auxiliary premises). The two theses are thus logically independent.

There are, moreover, important differences between incompatibilism* and the restricted versions of incompatibilism that the consequence argument is sometimes said to support. For one thing, those restricted versions of incompatibilism concern the relationship between determinism and various claims about free will. Incompatibilism*, by contrast, concerns the relationship between free will and necessitation by factors beyond the agent's control. The relata with which the theses are concerned are therefore different.[9] Note also that incompatibilism* isn't restricted to agents for whom there is a distant past or to those that aren't coy miracle workers. It applies to all agents, regardless of whether they have a past or can work miracles. In this respect, incompatibilism* is broader in scope than the various restricted versions of incompatibilism.

Incompatibilism* also suggests a reason for thinking that those restricted incompatibilist theses are true. Necessarily, for any possible world w containing agents relevantly like us (i.e., agents with a distant past and that aren't coy miracle workers), if determinism is true at w, then the behavior of the agents in question is necessitated by factors beyond their control, namely, events in the distant past together with the laws of nature. But, necessarily, if everything those agents do is necessitated by factors beyond their control, then they don't have free will (this

---

[8] The same can't be said about the restricted versions of incompatibilism mentioned earlier, both of which are entailed by (though they don't entail) incompatibilism. Necessarily, if determinism is incompatible with free will, then it's incompatible with free will in agents who are like us in relevant respects.

[9] This is clear from the fact that our behavior might be necessitated by circumstances beyond our control even if determinism is false. And determinism arguably doesn't entail that everything people do is necessitated by circumstances beyond their control. The thesis of determinism and the hypothesis that people's behavior is necessitated by circumstances beyond their control therefore don't come to the same thing.

follows from incompatibilism*). So, necessarily, for any possible world *w* containing agents relevantly like us, if determinism is true at *w*, then those agents lack free will.

Finally, recall Cutter's (2017, p. 279) suggestion that the conclusion of the consequence argument merits the name 'incompatibilism' only if it "would not be endorsed by paradigmatic compatibilists." Incompatibilism* satisfies this constraint, even if, as Cutter contends, the restricted versions of traditional incompatibilism don't. Most compatibilist would insist, contrary to incompatibilism*, that it is possible for agents to have free will even if everything those agents do is necessitated by factors beyond their control. A compelling argument for incompatibilism* would thus advance the debate between incompatibilists and (many of) their opponents.

Indeed, incompatibilism* is a thesis that many philosophers working on the topic of free will have had in mind all along.[10] There is, moreover, reason to think that it's actually the more central incompatibilist position. Consider, for instance, the fact that libertarians and other incompatibilists will often claim that it's possible for an agent to do something of his own free will even if what he did was determined, provided it was determined in an appropriate way by a prior exercise of free will on the part of that agent. This would be an exceedingly awkward claim if these philosophers really thought that an action's being determined *tout court* is incompatible with its being an exercise of free will on the part of its agent. The claim makes perfect sense, however, if what they really think is that an action's being determined entirely by factors beyond the agent's control is incompatible with its being an exercise of free will on the part of its agent.

Consider too the fact that no one, not even the staunchest incompatibilist, seems at all worried about future-to-past determinism. Standard definitions of determinism imply that our

---

[10] It's how Pereboom 2014 characterizes his own incompatibilist view, for example. It's also how Sartorio (2015, 2016) suggests that the incompatibilist position should be understood, and she contrasts her own compatibilist view of free will with something along the lines of incompatibilism*.

behavior is determined not only by facts about the past and the laws but by facts about the future and the laws as well. No one, however, is tempted to regard *this* aspect of determinism as problematic vis-à-vis free will. But why not? Why might it be tempting to think that past-to-future determinism is a threat to free will but not future-to-past determinism? As Sartorio (2016, p. 2) points out, the obvious answer is that the future seems to be within our causal reach, and thus within our control, in a way the past isn't. This suggests, in turn, that the reason why it can be tempting to see (past-to-future) determinism as a threat to free will is because it appears to entail the further deterministic hypothesis that our behavior is settled in advance by circumstances beyond our control, such as events in the distant past and the laws of nature. Now, whether determinism actually entails this hypothesis is debatable, for, as we've seen, it's controversial whether determinism entails that there is a remote past, and controversial as well whether laws of nature are necessarily beyond our control. The point, though, is that it's determinism's apparent connection to the further deterministic hypothesis just mentioned that can make it tempting to suppose that determinism is inimical to free will. But if that's right, the deeper question we should be asking is whether that further deterministic hypothesis really is inconsistent with free will. The real question, in other words, is whether incompatibilism* is true.

There is, then, good reason to think that incompatibilism* captures an important thesis about free will that has been lurking in the background all along and which is, in certain respects, a more fundamental incompatibilist position than the traditional incompatibilist thesis that free will and determinism are mutually exclusive. And while that traditional incompatibilist thesis may or may not find support from consequence-style arguments, incompatibilism* most certainly does. To see this, let *A* be any action of any agent *S*, suppose that *A* is necessitated by some set of circumstances, *X*, over which *S* has no control (call this the action-deterministic

hypothesis), and let "N$_S$$p$" abbreviate "$p$, and S has no choice, and never has had a choice, about whether p." Here are agent-relative versions of rules α and β:

(α*) $\Box p \vdash N_S p$.

(β*) $N_S (p \supset q), N_S p \vdash N_S q$.

I now offer the following consequence-style argument for incompatibilism*:

(1) $\Box (X \supset A)$	from the action-deterministic hypothesis
(2) $N_S X$	from the action-deterministic hypothesis
(3) $N_S (X \supset A)$	from 1 and rule α*
(4) $N_S A$	from 2, 3, and rule β*

The conclusion of this argument is that S has no choice, and never has had a choice, about whether he does A, and since S can be any agent in any possible world and A any action of that agent such that A is necessitated by factors beyond S's control, the conclusion generalizes: necessarily, no agent has, or ever had, a choice about what he does or doesn't do if his behavior is the inexorable consequence of circumstances beyond the agent's control.[11]

    Notice that, unlike the Third Formal Argument, this argument doesn't make the mistake of appealing to contingently true premises in the course of trying to establish a strict conditional by means of conditional proof. This argument therefore doesn't face the sorts of difficulties described above that seem to afflict other consequence-style arguments. I won't say more beyond

---

[11] An even simpler version of the argument can be obtained by replacing rules α* and β* with the following rule: $N_S p, \Box (p \supset q) \vdash N_S q$. This rule, in conjunction with (1) and (2), entails (4).

that about whether the argument is sound. (It depends entirely on whether rules α* and β* are valid.) The important point for present purposes is that, if sound, it establishes incompatibilism*, which, as I've argued, is a philosophically interesting thesis and is arguably a more central incompatibilist position than the one the consequence argument was initially designed to support.

References

Bailey, A. 2012. Incompatibilism and the past. *Philosophy and Phenomenological Research* 85: 351-376.

Beebee, H. & A.R. Mele. 2002. Humean compatibilism. *Mind* 111: 201-223.

Brueckner, A. 2008. Retooling the consequence argument. *Analysis* 68: 10-13.

Campbell, J. 2007. Free will and the necessity of the past. *Analysis* 67: 105-111.

Campbell, J. 2008. Reply to Brueckner. *Analysis* 68: 264-269.

Campbell, J. 2010. Incompatibilism and fatalism: reply to Loss. *Analysis* 70: 71-76.

Cutter, B. 2017. What is the consequence argument an argument for? *Analysis* 77: 278-287.

Finch, A. 2013. On behalf of the consequence argument: time, modality, and the nature of free action. *Philosophical Studies* 163: 151-170.

Huemer, M. 2000. Van Inwagen's consequence argument. *Philosophical Review* 109: 525-544.

Lewis, D. 1981. Are we free to break the laws? *Theoria* 47: 113-121.

Loss, R. 2009. Free will and the necessity of the present. *Analysis* 69: 63-69.

Loss, R. 2010. Fatalism and the necessity of the present: reply to Campbell. *Analysis* 70: 76-78.

McKay, T. & D. Johnson. 1996. A reconsideration of an argument against compatibilism. *Philosophical Topics* 24: 113-122.

Pereboom, D. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.

Rosen, G. 2002. The case for incompatibilism. *Philosophy and Phenomenological Research* 64: 699-706.

Sartorio, C. 2015. The problem of determinism and free will is not the problem of determinism and free will. In *Surrounding Free Will*, ed. A. Mele, 255-273. New York: Oxford University Press.

Shabo, S. 2011. What must a proof of incompatibilism prove? *Philosophical Studies* 154: 361-371.

Speak, D. 2011. The consequence argument revisited. In *The Oxford Handbook of Free Will* 2nd edition, ed. Robert Kane, 115-130. New York: Oxford University Press.

Stone, J. 1998. Free will as a gift from god: a new compatibilism. *Philosophical Studies* 92: 257-281.

van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Warfield, T. 2000. Causal determinism and human freedom are incompatible: a new argument for incompatibilism. *Philosophical Perspectives* 14: 167-180.

Widerker, D. 1987. On an argument for incompatibilism. *Analysis* 47: 37-41.