

Free Will, Resiliency and Flip-flopping

James Cain

Oklahoma State University

Abstract: Many philosophers accept with certainty that we are morally responsible but take it to be an open question whether determinism holds. They treat determinism as epistemically compatible with responsibility. Should one who accepts this form of epistemic compatibilism also hold that determinism is metaphysically compatible with responsibility—that it is metaphysically possible for determinism and responsibility to coexist? John Martin Fischer gives two arguments that appear to favor an affirmative answer to this question. He argues that accounts of responsibility, such as his, that are neutral with respect to whether responsible actions are determined have a "resiliency" that counts in their favor. Furthermore, he criticizes libertarians who argue on *a priori* grounds that determinism cannot coexist with responsibility and who admit that they would retract their argument if determinism were shown to hold; this "metaphysical flip-flopping" is said to render their positions implausible. I assess the merits of these arguments.

I. Introduction

Many philosophers consider it a certainty that we are morally responsible and consider it an open question whether determinism holds.¹ For them determinism is, in a sense, epistemically compatible with the existence of moral responsibility—that is, it is in some sense epistemically

possible that determinism holds and we are morally responsible. For the purposes of our discussion, I will assume that this is a reasonable position to take. The question then arises: should the considerations that lead one to this form of epistemic compatibilism also lead one to accept the metaphysical compatibility of determinism and responsibility—i.e., should it lead us to accept that it metaphysically possible for responsibility to exist in a deterministic world?²

John Martin Fischer develops arguments that appear to favor an affirmative answer to this question. He argues that accounts of responsibility that are neutral with respect to the question of whether responsible actions are determined have a "resiliency" that counts in their favor—they do not leave our status as responsible agents "hanging on a thread" should it turn out that determinism holds. Furthermore, he argues that if one were to both (a) endorse on *a priori* grounds a theory that denies the compatibility of responsibility and determinism and (b) admit that one would abandon the theory were determinism shown to hold, then (even if determinism was not shown to hold) one would be subject to a charge of "metaphysical flip-flopping" that would render implausible one's endorsement of the theory. I will argue that these considerations of Fischer do not give us reason to prefer theories that treat determinism as metaphysically compatible with responsibility.

There are special cases in which considerations that lead one to accept epistemic compatibilism should also lead one to accept metaphysical compatibilism—or at least to accept that it is highly likely that metaphysical compatibilism holds. Since the actual truth of a proposition entails its metaphysical possibility, evidence that a conjunction (X & Y) is highly likely supports both the view that X and Y are epistemically compatible and the view that they are metaphysically compatible.

Suppose on the other hand that one has evidence that there is at least a small likelihood that both X and Y are true. This gives us reason to hold that X and Y are epistemically compatible, but does it give us evidence that X and Y are metaphysically compatible? Since actuality implies metaphysical possibility, a small likelihood that X and Y both hold brings with it at least a small likelihood that X and Y are metaphysically compatible. But it need not do more than that. Suppose, for example, that at some point in history there was, for a given epistemic agent, certainty that bats exist, a 90% likelihood that to be a bat is to be a kind of mammal, a 10% likelihood that to be a bat is to be a kind of bird, and certainty that it is metaphysically impossible for a mammal to be a bird. For that person, something's being a bat was epistemically compatible with its being a bird. Nonetheless, under the circumstances it would have been reasonable for her to hold that, most likely, being a bat is metaphysically incompatible with being a bird.

Fischer has developed a couple of arguments that may seem to show that, even if one thinks that there is a low likelihood that determinism holds, a firm commitment to the epistemic compatibility of determinism and responsibility carries with it a strong (though perhaps defeasible) reason to accept their metaphysical compatibility. Unfortunately, in his writings Fischer does not, as far as I can see, explicitly distinguish between epistemic and metaphysical compatibility. So, without trying to nail down the precise sense of "compatibility" that Fischer has in mind, I will investigate whether his arguments can be taken in a way that provides a special reason to think that a commitment to the epistemic compatibility of determinism and responsibility carries with it a commitment to their metaphysical compatibility.

II. The argument from resiliency

Fischer has worked out a careful metaphysical account of free and responsible action which he refers to as "semicompatibilism". Though semicompatibilism allows that causal determinism is compatible with free and responsible action—and is thus a form of compatibilism—it does not require that determinism be compatible with agents' having freedom to do otherwise. I will not be concerned with the details of Fischer's theory; rather, my focus will be on how Fischer thinks compatibilism figures into the desiderata of a theory of responsibility. Fischer (2012, p. 140) remarks:

It counts in favor of my Semimcompatibilism (sic) that our status as morally responsible and as persons does not "hang on a thread;" we would not have to reconfigure our fundamental view of ourselves (at least in central aspects) or our basic metaphysical doctrines if we were to be convinced that causal determinism obtains. ... If we were to wake up to the *New York Times* headline, "The Natural Laws Have Associated With Them 100 % Probabilities," would it really be appropriate fundamentally to change our conception of ourselves—to give up on moral responsibility and personhood?

Let us grant, for the sake of argument, that determinism is epistemically compatible with responsibility. Furthermore, let us grant, in agreement with Fischer, that we should give a negative answer to his question: The mere discovery that determinism is true should not lead us to abandon our understanding of ourselves as morally responsible. Fischer counts it as a point in favor of his semicompatibilist theory that, were it discovered that determinism is true, he could give this answer without changing his account of responsibility. Fischer (2012, p. 118) says it

gives his theory "resiliency" that "our status as morally responsible does not 'hang on a thread'" depending on the outcome of "subtle ruminations of theoretical physicists".

I do not think that the resiliency of accounts like Fischer's in the face of hypothetical evidence for determinism should be seen as a reason to accept the metaphysical compatibility of determinism and responsibility. To see why, consider the following example. In the early days of the development of chemistry, a proponent of the view that water is H₂O (and thus water's existence is metaphysically incompatible with the nonexistence of H₂O) might have lingering doubts—she might worry that scientists will prove that there's no such thing as H₂O. Would she feel that the world's water supply is "hanging on a tread"? No, the existence of water is not in question. All that hangs by a tread is the verdict with respect to the correctness of her theory of the nature of water. The mere epistemic possibility that science will disconfirm the existence of H₂O does not give her a reason to hold that the existence of water is metaphysically compatible with the nonexistence of H₂O. Or, to be more precise, since actuality implies metaphysical possibility, the fact that there is some likelihood that water is not H₂O implies that there is some likelihood that being water is metaphysically compatible with being something other than H₂O. But there need not be a greater likelihood to the claim that being water is metaphysically compatible with not being H₂O than there is to the claim that water is not H₂O, and the likelihood of the latter claim might be very small.

Similarly, should science prove that determinism holds, there might be reason to reject particular accounts of moral agency that entail metaphysical incompatibilism. But the mere epistemic possibility that science *might* prove that determinism holds does not give us a reason to maintain that incompatibilist theories of responsible agency are in fact false. If there is a small

likelihood that determinism holds that need not give us more than a small likelihood that determinism is metaphysically compatible with moral responsibility.

II. The argument from flip-flopping

Let us turn to Fischer's second argument that could be taken to support the move from epistemic to metaphysical compatibilism. Fischer claims that there are special considerations that apply to libertarians like Peter van Inwagen who argue on *a priori* grounds that determinism is incompatible with free and responsible action and yet say that, if determinism were shown to be true, they would change their position and hold that free and responsible action turns out to be compatible with determinism after all.³ Though Fischer focuses on van Inwagen, I will try to address Fischer's argument without going into the details of van Inwagen's position. Fischer bases his argument on two features he claims to find in van Inwagen's presentation. They are:

(1) Van Inwagen constructs an argument for the incompatibility of determinism and moral responsibility that crucially appeals to a principle (which, following Fischer, we will refer to as *the Principle*) that is accepted *a priori*. The Principle (in conjunction with the claim that we are responsible and some other, acceptable claims) entails the falsity of determinism.

(2) Van Inwagen admits that, if determinism were conclusively shown to be true, he would give up the Principle and accept that determinism is compatible with the existence of moral responsibility.⁴

Fischer (2016, pp. 48-49) calls "the rejection of an *a priori* ingredient in the incompatibilist's argument, contingent upon learning that causal determinism is true, 'metaphysical flip-flopping'", and he holds that "van Inwagen's approach is implausible insofar as he is open to such metaphysical flip-flopping".

It might seem that what Fischer calls "flip-flopping" is merely a case of changing one's mind in the face of counterevidence. But Fischer thinks that a disposition to engage in metaphysical flip-flopping shows a weakness in one's position even if the counterevidence never materializes. Fischer reasons as follows: Van Inwagen admits that if scientists demonstrated that determinism holds he would give up the Principle; he even says:

...it would not surprise me too much to find that [the Principle,] which at present seems to me to be a truth of reason, had been refuted by the progress of science. (van Inwagen 1983, p. 221, quoted in Fischer, 2016, p. 51)

But, Fischer (2016, p. 56) argues:

...if there is a reason to reject the Principle in the counterfactual scenario where physicists report that determinism is true, that reason is already present in the actual scenario. Thus flip-flopping leads from libertarianism to compatibilism because it requires that there would be, and *is*, reason to reject the Principle.

We may take van Inwagen to be committed to a form of epistemic compatibilism: for all van Inwagen knows for certain, it might be the case that determinism holds and yet we are free and responsible. The question facing us is whether Fischer's argument shows that one holding this form of epistemic compatibilism already has reason to reject the *a priori* Principle and to accept the metaphysical compatibility of determinism and responsibility. I do not think so. A parallel example may help us to see why. Say that a collection of tiles has property *P* provided that the tiles are laid out in a rectangular arrangement with more than one row, where each row contains the same number of tiles, a number greater than one. I observe that there is a large collection of tiles before me with property *P*. Without counting the tiles, I reason my way to a conclusion, namely: "There are not 421 tiles in this collection." My reasoning crucially depends on a mathematical principle (which I call *the Mathematical Principle*) which I accept on *a priori* grounds, namely, that 421 is a prime number. I am very confident that this mathematical principle is true. Several times I have thought through justifications for the principle, and I can now see quite clearly that it must be true. And yet "...it would not surprise me too much to find that [the Mathematical Principle,] which at present seems to me to be a truth of reason, had been refuted." After all, I do make mistakes in my mathematical reasoning from time to time. I may imagine the epistemic possibility that people make an accurate count of the tiles and find that there are indeed 421 tiles. In the envisioned circumstance, I should think that I would flip-flop and say that I must have made a mistake, that I was wrong in my *a priori* reasoning. I would retract the claim that 421 is prime. And I expect that were that to happen—were I to find myself in the counterfactual scenario in which it was verified that there is a collection of 421 tiles with property *P*—I would be able to find a flaw in the reasoning that lead me to hold that 421 is prime. So I admit that I am disposed to flip-flop in this case.

Now let us see how we would assess this situation if we apply to it reasoning similar to that which Fischer applied to van Inwagen's admission of a disposition to flip-flop. As we saw, Fischer says with regard to van Inwagen, "...if there is a reason to reject the Principle in the counterfactual scenario where physicists report that determinism is true, that reason is already present in the actual scenario," and he concludes that flip-flopping leads from libertarianism to compatibilism. How ought we to respond if one were to say, "...if there is a reason to reject the Mathematical Principle in the counterfactual scenario where an accurate count finds there to be 421 tiles in a collection with property *P*, that reason is already present in the actual scenario"? Should we accept this and conclude that having property *P* is compatible with being a collection of 421 tiles?

In the counterfactual scenario there is a reason to reject the Mathematical Principle. In that scenario we are given reason to believe that there are 421 tiles arranged in a rectangle with, for some integers *m* and *n* both greater than one, *m* distinct rows each containing *n* tiles. That would give us reason to believe that, for such a pair *m* and *n*, $421 = m \times n$ and thus 421 would not be a prime number. So in that alternative scenario we would have reason to reject the Mathematical Principle. But notice that we would be unwarranted in saying, "if there is a reason to reject the [Mathematical] Principle in the counterfactual scenario ... that reason is already present in the actual scenario." In the actual scenario there is no reason to think that there are numbers greater than one that can be multiplied together to get 421. My disposition to flip-flop—when faced with strong evidence that there is a collection of 421 tiles that has property *P*—gives me no reason to reject the Mathematical Principle in the absence of such evidence. Furthermore, it gives me no reason to think that a collection of tiles' having property *P* is metaphysically compatible with its having 421 members.

Now let us return to Fischer's treatment of metaphysical flip-flopping. On a charitable reading of van Inwagen's position, he is willing to concede that, if in the future he were to find himself in the epistemically possible situation in which science makes it evident that determinism holds, he would come to accept both that determinism holds and that there is freedom and responsibility. If he were to grant that this epistemic possibility is also a metaphysical possibility, then of course that would commit him to accepting the metaphysical compatibility of determinism with freedom and responsibility. But it is unclear why he should grant this, and thus it is unclear that a disposition to metaphysically flip-flop gives one a strong reason to embrace metaphysical compatibilism. One inclined towards metaphysical flip-flopping in the way criticized by Fischer might well reason as follows:

I am convinced that there are things I have done for which I am responsible, and I think that I should continue to hold this belief even in the unlikely event that determinism is found to be true. If determinism actually holds then there is a metaphysically possible world—the actual world—in which there is both determinism and responsibility. Thus, if determinism actually holds, responsibility is metaphysically compatible with determinism. But if determinism does not actually hold this particular ground for accepting metaphysical compatibilism is no longer available: this reason to reject my current views is not "already present in the actual scenario". Furthermore, counterfactual epistemic scenarios need not be metaphysically possible, and so epistemic compatibility need not be taken as evidence for metaphysical compatibility.⁵

Works Cited

Cain, James. (2004) Free Will and the Problem of Evil. *Religious Studies* 40: 437-56.

Fischer, John Martin. (2012) Semicompatibilism and its Rivals. *Journal of Ethics* 16: 117-43.

Fischer, John Martin. (2016) Libertarianism and the Problem of Flip-flopping. In Daniel Speak and Kevin Timpe (eds), *Free Will and Theism: Connections, Contingencies, and Concerns* (pp. 48-61). Oxford: Oxford University Press.

van Inwagen, Peter. (1983) *An Essay on Free Will*. Oxford: Oxford University Press.

¹ Or at least they hold it to be an open question whether some form of "near determinism" holds. The arguments concerning determinism made in this paper apply as well to "near determinism", but to simplify the discussion I will focus on determinism. Here I am following van Inwagen's (1983, p. 198) use of the term "near determinism."

² I will not spell out in detail the notions of epistemic and metaphysical compatibility. However one demarcates a sense of possibility, there will be a corresponding sense of compatibility. So, if there are a variety of subtypes of metaphysical or epistemic possibility, there will be a corresponding variety of subtypes of compatibility. See my (2004) for a discussion of varieties of compatibilism.

³ Fischer raises this concern in several of his writings. I will concentrate on his 2016.

⁴ See van Inwagen 1983, p. 221. The principle in question is called "(β)" by van Inwagen and "the Principle of the Transfer of Powerlessness" by Fischer. Though Fischer (2016, p. 52) says van Inwagen "has written that, if he were to be convinced of the truth of causal determinism, he would give up the Principle of the Transfer of Powerlessness and embrace compatibilism," van Inwagen (1983, pp. 219-21) actually makes a more qualified claim indicating that he would most likely give up (β) if science were to demonstrate that determinism holds.

⁵ For thoughtful comments on this paper, I would like to thank Rebecca Bensen Cain, Matthew Shea, and J.P.

Andrew.