7

Symposium on the Fixity of the Past

7.1

Incompatibilism and the Fixity of the Past

Neal A. Tognazzini and John Martin Fischer

7.1.1

A style of argument that calls into question our freedom (in the sense that involves freedom to do otherwise) has been around for millennia; it can be traced back to Origen. The argument-form makes use of the crucial idea that the past is over-and-done-with and thus fixed; we cannot now do anything about the distant past (or, for that matter, the recent past)—it is now too late. In ancient and medieval times, the argument was in service of "fatalism"—logical and theological. That is, the argument purported to show that prior truth-values of statements about future human behavior make it the case that the behavior is not free, or that God's prior beliefs about future human behavior similarly make it the case that the behavior is not free.

In the modern era, with the rise of science, a new way of filling in the venerable argument-template has emerged, issuing in what Peter van Inwagen has called "The Consequence Argument".[1] Van Inwagen gave it this name because, if causal determinism is true, then all our behavior is the consequence of the past and laws of nature. More specifically, if causal determinism is true, statements about the past (intrinsically construed—the temporally nonrelational past), together with the laws of nature, entail all truths about present and future human behavior. It can thus seem mysterious how, if causal determinism were true, human beings could be free. It can

[1] Van Inwagen (1983: 16).

thus also seem unclear that we can legitimately be deemed morally responsible for any of our behavior.

Various contemporary philosophers have sharpened these worries into an argument (with the same general form as the ancient argument stemming from the fixity of the past).[2] Peter van Inwagen's great book, *An Essay on Free Will*, built on the work of these philosophers by presenting and defending the argument in perhaps the clearest and most forceful way ever. It has thus been an enormously important and influential work, for which we have great admiration.

This is not to say that all philosophers have rushed to accept van Inwagen's view that the Consequence Argument is sound. Some philosophers have resisted its validity, whereas others have called into question its soundness.[3] In fact, there is a strong argument to be made that the debate over whether determinism rules out the ability to do otherwise is at a stalemate.[4] Incompatibilists think that evaluating claims about what agents can do requires holding fixed the past and the laws of nature; compatibilists disagree. In particular, whereas the incompatibilist argument is motivated crucially by appeal to two "fixity" principles—the Principle of the Fixity of the Laws and the Principle of the Fixity of the Past—various compatibilists simply think we ought to reject one or the other of those principles. ("Multiple-pasts compatibilists" think we need not hold fixed the past when evaluating claims about what an agent can do, and "local miracle compatibilists" think we need not hold fixed the laws.) There appears to be little that either side can say to convince the other, and it is difficult to see how one might successfully *argue* for the relevant fixity principles that drive the incompatibilist argument.

Recently, however, in a bold new paper, Wes Holliday has attempted to break this seeming stalemate by presenting a new argument for the Principle of the Fixity of the Past.[5] Holliday's argument is subtle and ingenious, and worthy of serious consideration, especially given the promise it holds for genuinely advancing this old debate. In what follows, however, we argue that despite its considerable ingenuity, Holliday's argument fails to convince, and the stalemate appears to remain.

7.1.2

First, let us take a look at the relevant argument for incompatibilism, in which the Principle of the Fixity of the Past features so prominently. One version, stated informally, runs as follows:

1. If determinism is true, then the past and the laws of nature together entail every action that anyone ever actually performs.

[2] See, for instance: Ginet (1966: 87–104); and Wiggins (1973: 31–62).

[3] For developments and evaluations of such strategies, see, for example: Slote (1982: 5–24); Lewis (1981: 113–21); and Fischer (1994: Chapter 4).

[4] John Martin Fischer makes a detailed case for this conclusion in *The Metaphysics of Free Will*, and van Inwagen's own considered view is that it is utterly mysterious how any of us can possibly have free will, given the "seemingly unanswerable" arguments for its incompatibility with both determinism and indeterminism. See van Inwagen (2008: 327–41).

[5] Holliday (2012: 179–207).

2. So, if determinism is true, then if someone were to do anything other than what they actually do, either the past would have to be different or the laws of nature would have to be different.
3. But if, in order for someone to do otherwise, the past would have to be different, then that person cannot do otherwise.
4. And if, in order for someone to do otherwise, the laws would have to be different, then that person cannot do otherwise.
5. So, if determinism is true, then no one can perform any action other than the actions they actually perform.

Premise 3 is an informal statement of the Principle of the Fixity of the Past, and premise 4 is an informal statement of the Principle of the Fixity of the Laws. These two premises are the controversial premises in the argument, and they are at the heart of the alleged stalemate.

For the sake of his paper, Holliday presupposes that premise 4 is true, and instead focuses his attention on premise 3, the Principle of the Fixity of the Past, offering a new argument for it.[6] His argument—the "Action-Type Argument for the Principle of the Fixity of the Past" (p. 189)—relies crucially on introducing the action type *action that is inconsistent with the past*, which he proposes to understand as follows: an action such that if *s* were to do it, then the past would (have to) be different.[7] Given the connection between this type of action and the Principle of the Fixity of the Past, we might recast the latter principle for ease of exposition as follows:

(FP) An agent cannot perform an action that is inconsistent with the past.

Holliday's argument for (FP), in a nutshell, is this: (i) necessarily, no one ever does perform an action inconsistent with the past; (ii) if there is no world in which an action of type X is ever performed, then necessarily, no one *can* perform an action of type X; so, (iii) necessarily, no one can perform an action inconsistent with the past.[8] That is, Holliday attempts to argue for a claim about what agents *can* do from a claim about what it is *metaphysically possible* for agents to do. Since there is no possible world in which an agent *does* perform an action that is inconsistent with the past (the nonoccurrence of such actions is, after all, entailed by the past), no agent *can* perform such an action. Again: (i) actions that are inconsistent with the past are impossible (there is, after all, no possible world at which such an action is performed); (ii) no one can do the impossible; so, (iii) no one can do anything inconsistent with the past. And that's just to say that (FP) is true.

It will be instructive to consider a quick compatibilist reply to this argument, because it will bring out the need for an important clarification, which will then lead to a full-blown statement of Holliday's argument. The quick compatibilist reply is this:

[6] Since Holliday merely presupposes (rather than argues for) the truth of the Principle of the Fixity of the Laws, he acknowledges that his argument does not address so-called local miracle compatibilists, who reject this premise. See, for example, David Lewis, "Are We Free to Break the Laws?".

[7] Holliday (2012: 186–7). We have omitted the double time-indexing in Holliday's original explication of this action type because it is inessential for our purposes.

[8] Holliday (2012: 201).

We're willing to grant that impossibility entails inability, but we deny that it is impossible to perform an action that is inconsistent with the past. After all, actions that we are determined not to perform are typically such that, *had we wanted to perform them*, we would have. In other words: even if we are actually determined not to perform some action, there will still be plenty of worlds in which we *do* perform that action, and those worlds will simply be ones with a different past. So it's not impossible to perform an action inconsistent with the actual past; it's just that no one ever will perform such an action. And *will not* manifestly does *not* entail *cannot*.

This is a typical compatibilist move and Holliday anticipates it. In response, Holliday distinguishes between two senses in which an action might be said to be inconsistent with the past, depending on which "past" is at issue. As Holliday puts it, an action may either be of type *F* or of type *I*, where each type can be understood according to the following functions on world-time pairs (and where '$w_@$' is the name of the actual world):

F(w, t) = the set of actions inconsistent with the past relative to *t* of $w_@$.

I(w, t) = the set of actions inconsistent with the past relative to *t* of *w*.[9]

What the quick compatibilist reply gets right is that actions of type *F* are not thereby impossible: after all, just because an action is inconsistent with the past of the *actual world* does not mean that there is no world in which that action is performed; it is just that it will only be performed in worlds with pasts that differ in some way from the actual past. But what the quick compatibilist reply overlooks, according to Holliday, is that an action one is determined not to perform will also be an action of type *I*, and actions of type *I* are impossible: in no world does anyone perform an action that is inconsistent with the past *of that world*. The sense of 'inconsistent with the past' at issue in Holliday's argument, then, is the sense associated with actions of type *I*. Indeed, Holliday maintains that it is precisely because the difference between type *F* and type *I* actions has gone unnoticed that incompatibilists haven't previously noticed an argument for the Fixity of the Past to which they are entitled.

With this distinction in mind, we can informally restate Holliday's argument as follows: necessarily, no one ever performs an action of type *I*; if there is no world in which an action of type X is ever performed, then necessarily, no one *can* perform an action of type X; so, no one can perform an action of type *I*. That is: the Principle of the Fixity of the Past is true. And then the incompatibilist argument can continue as follows: actions we are determined not to perform are actions of type *I* (they fall within the range of the above function that specifies how we are to understand type-*I* actions); so, no one can perform an action that he is determined not to perform.

Holliday gives both an informal 'natural language' version of his argument, as well as a formal version using symbols, and the worries we plan to raise about the argument require that we have both versions in front of us. So, first, let us fix our symbols (following Holliday):

D(*s, y, w, t*) = agent *s* does action *y* in world *w* at time *t*. ('D' for 'does'.)

C(*s, y, w, t*) = agent *s* can, in world *w* at time *t*, do action *y*. ('C' for 'can'.)

[9] Holliday (2012: 191).

$X(w, t)$ = the set of all actions that fall under type X in world w relative to time t.

$y \in X(w, t)$ = action y falls under action type X in w relative to t.[10]

Now, we can state Holliday's two-premise argument for the Principle of the Fixity of the Past as follows (p. 201):

(1) An agent cannot perform an action of type X if there is no possible world in which an agent performs an action of type X.

$\forall X\,[\neg\exists s, w, t, y\,(y \in X(w, t) \wedge \mathrm{D}(s, y, w, t)) \rightarrow \forall s, w, t, y\,(y \in X(w, t) \rightarrow \neg\mathrm{C}(s, y, w, t))]$

(Read from the symbols: for all action types, if there's no world-time pair at which anyone *does* perform an action which is of a certain type at that world-time pair, then there's no world-time pair at which anyone *can* perform an action which is of that type at that world-time pair.)

(2) There is no possible world in which an agent performs an action that is inconsistent with the past (an action of type I).

$\neg\exists s, w, t, y\,(y \in I(w, t) \wedge \mathrm{D}(s, y, w, t))$

(Read from the symbols: there is no world-time pair at which any agent ever does perform an action which is type I at that world-time pair.)

(3) Therefore, an agent cannot perform an action that is inconsistent with the past (an action of type I).

$\forall s, w, t, y\,(y \in I(w, t) \rightarrow \neg\mathrm{C}(s, y, w, t))$

(Read from the symbols: for every agent, action, and world-time pair, if the action is type-I at that world-time pair, then the agent cannot perform it.)

The argument looks complicated, but in fact it is strikingly straightforward and elegant: (1) impossibility implies inability; (2) type-I actions are impossible; so, (3) no one is able to perform type-I actions (i.e. actions that are inconsistent with the past). The argument is *so* straightforward, in fact, that it is difficult to see how it might be resisted. Nevertheless, we think it ought to be resisted, and in the remainder of the paper we explain how.

7.1.3

We actually have two related worries about the argument, and both are directed at premise (1). The first worry is that premise (1) begs the question against the compatibilist because it presupposes an understanding of 'can' that the compatibilist has antecedent reason to reject. The second, related worry is that premise (1) does not in fact adequately capture the intuitive idea that no one can do what is impossible, and hence is unmotivated. We'll begin with the charge of begging the question.

To begin, note that premise (1) is meant to be a generalized statement of the seemingly uncontroversial thesis that if some action is impossible, then no one can

[10] We have simplified Holliday's formalization somewhat, but in ways that are inessential to the points we wish to make.

perform it. What makes it generalized is that it is stated not in terms of particular actions but in terms of action *types*. So, at the level of action types, the claim is that if some action falls under an impossible action type, then no one can perform actions of that type, where an action type is impossible just in case there is no world in which anyone ever performs an action of that type. But when we generalize the claim in this way, we introduce an ambiguity. When we say that there is no world in which anyone ever performs an action of a certain type, we might mean:

(a) There is no world in which anyone ever performs an action which is, *in that world*, of type X.

$\neg\exists s, w, t, y\ (y \in X(w, t) \wedge \mathrm{D}(s, y, w, t))$

or we might mean:

(b) There is no world in which anyone ever performs an action which is, *as a matter of actual fact*, of type X.

$\neg\exists s, w, t, y\ (y \in X(w_{@}, t) \wedge \mathrm{D}(s, y, w, t))$

We are not claiming that Holliday's first premise is ambiguous, though, because his formalization makes it clear that he intends reading (a). The formalized version of premise (1), again, is this:

(1) $\forall X\ [\neg\exists s, w, t, y\ (y \in X(w, t) \wedge \mathrm{D}(s, y, w, t)) \rightarrow \forall s, w, t, y\ (y \in X(w, t) \rightarrow \neg\mathrm{C}(s, y, w, t))]$

His first premise, then, is a (universally quantified) conditional, which moves from the claim that there is no world in which anyone *does* perform an action which is, in that world, of type X, to the claim that there is no world in which anyone *can* perform an action which is, in that world, of type X.

Now, we said earlier that we think this premise presupposes a question-begging sense of 'can', and this is most easily seen by looking at its contrapositive, which gives a necessary condition on an agent's ability to perform an action:

(1C) $\forall X\ [\exists s, w, t, y\ (y \in X(w, t) \wedge \mathrm{C}(s, y, w, t)) \rightarrow (\exists s, w, t, y\ (y \in X(w, t) \wedge \mathrm{D}(s, y, w, t))]$.

In other words:

(1C) An agent *can*, at some world-time pair, perform an action that is of type X at that world-time pair only if there is some world-time pair at which an agent *does* perform an action which is, at that world-time pair, of type X.

At first blush, this may seem to say no more than the innocuous and utterly uncontroversial claim that an agent is only able to do those things that it is possible to do. But in fact it says much more than that. In particular, it says that an agent is able to perform an action which falls under a certain type only if there is a possible world in which some agent performs an action *of the same type*. It is this restriction—that an actual-world ability to perform a certain type of action requires an other-worldly performance of an action *of the same type*—that causes trouble for Holliday's

argument. Just as the compatibilist will deny that we need to hold fixed the past and the laws when evaluating which other-worldly performances are relevant to actual-world ability claims, so will they deny that we need to hold fixed the action type.

To see the problem with Holliday's argument more clearly, consider a simple-minded compatibilist view, according to which an agent *can* perform an action if and only if the agent *would* perform the action *were* he to desire to perform it.[11] Now suppose, to take Holliday's example, that in the actual world Themistocles is determined (i.e. causally determined) not to send his fleet to Corinth. The question is: despite his being determined not to choose Corinth, *can* he? Holliday says no, since choosing Corinth is inconsistent with the past, and in no world does anyone perform an action that is inconsistent with the past. But our simple-minded compatibilist will be puzzled why there needs to be such a world in order for Themistocles to be able to choose Corinth. After all, on this compatibilist's view, so long as there is some sphere of nearby worlds in which Themistocles both desires to choose Corinth and does choose Corinth, that will be sufficient for the truth of the claim that Themistocles *can* choose Corinth. True, the compatibilist will admit, in those nearby worlds, choosing Corinth is not an action that is inconsistent with the past (those worlds have different pasts), so choosing Corinth is not, in those worlds, type *I*. It is, of course, type *I* in the actual world, but our compatibilist will not be bothered by that, for his analysis of 'can' does not require that we hold fixed the type of action when evaluating claims about what agents can do.

In his attempt to motivate premise (1), Holliday points out that when it is true that an agent can perform an action, "there should be some possible world that 'witnesses' the truth of this can-claim" (p. 194). But our envisaged compatibilist does not deny this; rather, the issue is whether the witnessing world needs to be one at which the action is *of the same type* as it is in the world of the can-claim. Holliday's premise (1) requires that it is; but the compatibilist will maintain that this requirement stacks the deck against him. *Of course* there is no world in which an action is performed which is, *in that world*, inconsistent with the past. But all the compatibilist requires is that there be some suitable world in which an action is performed which is inconsistent with the past *of our world* (thus 'witnessing' the truth of the actual-world can-claim despite the truth of determinism). And this means that our "quick compatibilist reply" from above is actually, suitably embellished, the right reply to make to Holliday's argument: just because an action is inconsistent with the past of the actual world does not mean that there is no world in which that action is performed, and just because an action falls under an impossible action type does not mean that there is no world in which that action is performed; it's just that such an action will only be performed in worlds where it doesn't fall under that type. And the existence of at least some of these worlds will be sufficient for the truth of the claim that an agent *can* perform the action despite his being determined not to.

This brings us to our second, related worry about Holliday's first premise, which is that it is not adequately supported by the intuitive claim that no one can do what is

[11] This simple conditional analysis of 'can' is problematic for well-known reasons, but its simplicity makes it well-suited for illustrating our point, which is a point about compatibilist views of 'can' in general, even more sophisticated ones.

impossible. Formulated in the way it needs to be for Holliday's "Action-Type Argument" to be valid, it says that no one can do things that fall under action types that are necessarily uninstantiated. But to say that some action falls under an action type that is necessarily uninstantiated is much different than saying that the action is *impossible*. For although the type "inconsistent with the past" is indeed necessarily uninstantiated, many actions that fall under this type are nevertheless possibly performed; it is simply that in those worlds in which they are performed, they do not fall under that type.

We have said that our two worries are related, so let us briefly explain what we have in mind here. It is a tricky matter to determine whether an argument is question-begging, but in order to substantiate that charge, it is clearly not enough merely to point out that the argument contains some premise that an opponent of the conclusion will reject.[12] (If that were sufficient, then every valid argument would count as question-begging.) We have made that point about Holliday's argument—that the compatibilist will reject premise (1) as it is formulated—but we have also said something more, which is that premise (1) is given *no independent motivation*. It seems at first to be motivated by the claim that no one can do the impossible, but we have argued that this claim does not in fact support the much more robust premise (1). Without an independent motivation to accept premise (1), then, its role in the argument is no more than a bit of incompatibilist foot-stamping, and *this* is the sense in which it is dialectically infelicitous.

It is controversial, of course, whether the existence of a non-actual possible world in which some agent performs an action inconsistent with the actual past confers on the agent the ability, in the actual world, to perform some action that is inconsistent with the actual past. The compatibilist maintains that the existence of certain of those other worlds (with alternative pasts) means that agents in the actual world nevertheless *can* perform actions they are determined not to perform; the incompatibilist disagrees. But our point is simply that Holliday's argument does not offer any *new* reason for thinking we should go with the incompatibilist here. Instead, it simply crystallizes the incompatibilist intuition.[13]

7.1.4

The argument for incompatibilism that relies on the fixity of the past and the laws—the Consequence Argument—is highly contentious. Whereas many philosophers have accepted it as sound, others hold that it is unsound. Given the roster of distinguished philosophers who have vigorously disagreed about the argument, it would be a great intellectual contribution if someone could establish one of its key premises. In his bold and fascinating paper, Holliday offers a new argument for the fixity of the past premise. This premise may well be true, but we have argued that, despite his inventive and sophisticated argumentation, Holliday has not succeeded in establishing its truth.

[12] For an insightful discussion of begging the question, see Lecture 3 of van Inwagen (2006).

[13] Moreover, since our critique is aimed at premise (1) of Holliday's argument, the very same critique would apply to the "direct" argument for incompatibilism that he presents at the end of his paper, whose first premise is the same. See Holliday (2012: 202–5).

Even if Holliday's notable attempt is not entirely successful, we are all in debt to Peter van Inwagen for articulating the Consequence Argument in a clear and sharp way—a way that makes it possible to present and evaluate new and promising defenses of its premises. It is a great virtue of van Inwagen's regimentation of the ancient argument that it provides a framework within which the basic issues can be addressed in a fruitful way.

Acknowledgments

We are extremely grateful to Patrick Todd and John Keller for detailed comments on this paper, and for helping us to see how best to articulate the ideas presented here. Thanks also to the anonymous referees for their comments.

Bibliography

Fischer, J. M. (1994), *The Metaphysics of Free Will* (Blackwell), Chapter 4.

Ginet, C. (1966), 'Might We Have No Choice?', in K. Lehrer (ed.), *Freedom and Determinism* (Random House), 87–104.

Holliday, W. (2012), 'Freedom and the Fixity of the Past', in *The Philosophical Review* 121.

Lewis, D. (1981), 'Are We Free to Break the Laws?', in *Theoria* 47: 113–21.

Slote, M. (1982), 'Selective Necessity and the Free-Will Problem', in *Journal of Philosophy* 79: 5–24.

Van Inwagen, P. (1983), *An Essay on Free Will* (Clarendon Press.)

Van Inwagen, P. (2006), *The Problem of Evil* (Oxford University Press).

Van Inwagen, P. (2008), 'How to Think About the Problem of Free Will', in *The Journal of Ethics* 12: 327–41.

Wiggins, D. (1973), 'Toward a Reasonable Libertarianism', in T. Honderich (ed.), *Essays on Freedom of Action* (Routledge and Kegan Paul), 31–62.