

Freedom and the Fixity of the Past*

Wesley H. Holliday

Stanford University

1. Introduction

To find out whether a person could have done something, such as prevent a disaster or save a victim, we typically ask about the person's abilities and opportunities. Did the person know how to intervene? Was the person in a position to help? Few of us would bother to ask the question: what are the *laws of nature* that governed the situation? Yet a number of philosophers have argued that if deterministic laws govern human action, then it is easy to find out whether someone could have done something.¹ For the only things one could have done in life are the very things one did.

In a world with *deterministic* laws of nature, at any time there is only one future evolution of the world that is possible given the laws and the state of the world at that time. Determinism is the thesis that the laws of our world are deterministic. If determinism is true, then the laws of nature and the initial conditions of the Big Bang determined a unique future for our world; every movement you will ever make is part of that unique future, determined eons ago.²

* I am grateful to John Perry and John Martin Fischer for much valuable discussion and encouragement. I also thank Thomas Icard, Krista Lawlor, Neil van Leeuwen, Assaf Sharon, Jack Spencer, Patrick Todd, Neal Tognazzini, and the anonymous referees for the *Philosophical Review*, who provided many helpful comments.

¹ See, for example, Ginet 1966; van Inwagen 1975, 1983; Lamb 1977; and Wiggins 1973.

² This way of describing determinism assumes a view of the laws of nature according to which a law is not merely a regularity that turns out to be exceptionless over all time, but rather something that "governs" events as they unfold. I discuss this assumption in section 5.

In what follows, we will consider the traditional question of whether determinism is compatible with the “freedom to do otherwise,” leaving aside the question of whether determinism is true. Consider a historical example. In one of the final battles of the Greco-Persian war, the Athenian commander Themistocles had a choice to make: lure the Persian navy into straits near the island of Salamis or engage the Persians off the coast of Corinth.³ Themistocles chose Salamis, despite protests from his Peloponnesian allies. Could Themistocles have agreed with the Peloponnesians and chosen Corinth? If so, he had the *freedom to do otherwise* than order his fleet to Salamis. In general, let us say that an agent s had the freedom to do otherwise at a time t if and only if s could have done something at t other than what s actually did at t .⁴

If determinism is true, could Themistocles have ordered his fleet to Corinth instead of Salamis? Some philosophers, following Hume ([1777] 1993, sec. 8), would reply that determinism has nothing to do with what Themistocles could have done. What matters is that the commander of the Athenian navy presumably had the ability to order the fleet to Corinth and opportunities to do so. He simply decided not to. As far as we know, his decision was not subject to any constraint or compulsion, either physical or psychological, but was rather the product of his own deliberations. So what reason is there for thinking that Themistocles could not have ordered the fleet to Corinth? Had he wanted to, no laws of nature (deterministic or otherwise) would have prevented him from doing so.

³ For an account of the Battle of Salamis, see Strauss 2004. I take some liberties with the story for philosophical purposes.

⁴ In section 2, I adopt a practice of double time indexing, according to which we should say that at time t_* , agent s had the freedom to do otherwise at time t if and only if at t_* , s could have done something at t other than what s actually did at t .

Philosophers in the tradition of Hume argued along similar lines for the compatibility of Liberty and Necessity. More recent defenders of compatibilism have refined their position with sophisticated theories of how the freedom to do otherwise is possible in a deterministic world.⁵ According to these compatibilists, even if Themistocles was determined by the laws and the past to order the fleet to Salamis, all that follows is that he would not *exercise* his ability to order the fleet to Corinth, not that he lacked the ability.

Incompatibilists about freedom and determinism have a ready reply: Themistocles may have had a general ability to order the fleet wherever he pleased, but if he was determined by the laws and the past to order the fleet to Salamis, then he had no opportunity to do otherwise. What good are abilities with no opportunities to exercise them? So goes a traditional line of debate.

In this article, I present a new form of argument for incompatibilism.⁶ In section 2, I begin by discussing an argument that is not new, John Martin Fischer's Conditional Version of the Argument for Incompatibilism (1994, chap. 4). A key premise of the Conditional Argument is the *Principle of the Fixity of the Past*, a subject of much debate between compatibilists and incompatibilists. In section 3, I discuss why the debate appears to be stuck in what Fischer (1994, 83) calls a "Dialectical Stalemate." In section 4, I argue that the stalemate can be broken using a new Action-Type Argument for the Principle of the Fixity of the Past. Finally, in section 5, I show how the form of argument used to establish the Principle of the Fixity of the Past can be used to give a Simple Argument for incompatibilism, which bypasses the Conditional Argument.

⁵ See the references in Kane 2002, sec. 4.

⁶ I use the terms 'incompatibilism' and 'compatibilism' for views about the compatibility of determinism and the freedom to do otherwise, not about the compatibility of determinism and moral responsibility.

The conclusion is that incompatibilism follows on the assumption of a “governing” view of the laws of nature.

2. The Conditional Argument for Incompatibilism

The Conditional Argument involves thinking of our example of Themistocles and the fleet in a special way. About such a historical episode, it is natural to ask counterfactual questions. What if Themistocles had sent the fleet to Corinth instead of Salamis? How would future events have been different? Would Persia have conquered Greece, with far-reaching consequences to Western civilization? But suppose we ask a different question: If Themistocles had sent the fleet to Corinth instead of Salamis, how (if at all) would *previous events* leading to his decision have been different? Could everything have been just the same and yet Themistocles have sent the fleet to Corinth?

Imagine the situation at some time before the Battle of Salamis. If determinism is true, then it is determined by the laws of nature and the state of the world long before the moment of his decision that Themistocles will not order the fleet to Corinth. If Themistocles *were* to order the fleet to Corinth, then (ruling out miracles) the state of the world long before *would (have to) have been different*.⁷ But can Themistocles do something such that if he were to do it, the past would (have to) be different? He cannot, for the past is fixed. Therefore, if determinism is true, the only thing Themistocles can do is what he will do—order the fleet to Salamis.

⁷ Different authors express the conditional differently. Some write that if agent *s* were to do (or had done) action *y*, then the past would be (would have been) different. Others write that if *s* were to do (or had done) *y*, then the past would *have to* be (would *have to* have been) different. To satisfy both groups, I include the ‘have to’ in parentheses. Peacocke (1999, 326) argues that there is an important difference between the two conditionals. However, for the main argument of this article, whether we read the conditional with ‘have to’ does not matter.

From the reasoning above, we can extract the general Conditional Argument for incompatibilism, named after the *backtracking conditional*, ‘if he were to do it, the past would (have to) be different’. Moving in reverse order, the final premise in the argument is the Principle of the Fixity of the Past, which Fischer (1994, 78) states as follows:

For any action Y , agent S , and time t , if it is true that if S were to do Y at t , some fact about the past relative to t would not have been a fact, then S *cannot* at t do Y at t .

According to this principle, an agent cannot perform any action, the performance of which would require the past to have unfolded differently than it actually did. Let us rewrite the principle with a few modifications:

(FP) For any action y , agent s , and times t and t' ($t \leq t'$), if it is true that if s were to do y at t' , the past relative to t would (have to) be different, then s *cannot* at t do y at t' .

For action and agent variables, we will use lower-case italicized letters, reserving upper-case italicized letters for action type variables, to be introduced later. In place of Fischer’s expression ‘some fact about the past relative to t would not have been a fact’, we will use the shorter expression ‘the past relative to t would (have to) be different’, taking the phrase ‘the past relative to t' ’ to refer to *what happens* in the world up to t .

The reason for the additional time variable is to keep track of when an agent’s “window of opportunity” to perform an action closes. Imagine a Greek peasant farmer. At the age of fifteen, he may be able to become the commander of the Athenian navy at forty. However, at the age of thirty-nine, when he is still tending an olive orchard on the Greek mainland, he is no longer able to become the commander of the Athenian navy at forty. He can at fifteen become

the commander at forty, but he cannot at thirty-nine become the commander at forty.⁸ Hence the form: s can (cannot) at t do y at t' .

Returning to the case of Themistocles, let t' be the time at which he orders the fleet to Salamis. Let t be a year before. Important events that lead to Themistocles's decision to fight at Salamis occurred between t and t' . For example, the Athenians received a prophecy from Apollo's oracle at Delphi, which Themistocles interpreted to mean that they would defeat the Persians at Salamis. Given this fact, it may be that if Themistocles had ordered the fleet to Corinth at t' , the past *relative to* t would have been the same, while events *between* t and t' —such as those involving the oracle—would have been different. If this is the case, then for all we know, Themistocles could *at* t have ordered the fleet to Corinth at t' , since his doing so would not have required the past relative to t to be different. This last condition may not hold if determinism is true, but (FP) does not assume determinism.

If determinism is true and if Themistocles had ordered the fleet to Corinth instead of Salamis, how far back would the past have (to have) been different? At the least, if Themistocles had ordered the fleet to Corinth on September 25th, 480 BCE, perhaps he would not have set his famous trap on the night of the 24th, in which he sent one of his slaves with a message to the Persians, designed to lure them into the Salamis straits. In fact, we can reason that the past would have (to have) been different even farther back than that. For if determinism is true, there is only one history that is possible given the laws of nature and the initial conditions of the universe. In order for history to have unfolded differently in any respect, one of two things would have to

⁸ Beebee (2002, 239) gives an example of this kind, distinguishing times of ability/opportunity and the time of action.

have been the case. Either the initial conditions of the universe would have been different or there would have been a *miracle*, a violation of the actual laws of nature.⁹

In what follows, I will assume that none of our actual laws of nature would have been violated if agents had done otherwise in history.¹⁰ There would have been no miracles. Given determinism, this assumption is equivalent to the following *Backtracking Principle*:

(BT) For any action y , agent s , and times t and t' ($t \leq t'$), if s is *determined* not to do y at t' , then if s were to do y at t' , the past relative to t would (have to) be different.

Here ‘determined not to do y at t' ’ means that it is not possible for the agent to do y at t' , holding fixed the laws of nature and the initial conditions of the world.¹¹ It follows from (BT) that if determinism is true and if Themistocles had ordered the fleet to Corinth at t' , then the past would

⁹ I assume it is gratuitous to suppose both that the initial conditions would have been different and that there would have been a miracle in David Lewis’s sense of a violation of the actual laws of nature (Lewis 1979, 468-469). Of course, if there had been a violation of the actual laws, then there would have been a difference in the past, at least between the time of the law violation and the present. See n. 14 below.

¹⁰ See Bennett 1984 for an argument that no laws would have been violated, but rather the past would have been different back to any earlier time. This assumes we are not considering what would have happened if an agent had performed an action that is itself law-breaking, like running faster than the actual speed of light; for if an agent had done otherwise by doing that, then of course laws would have been violated.

¹¹ In section 5, we will use the notion of an agent’s being determined *at* t not to do y at t' , in the sense that it is not possible for the agent to do y at t' , holding fixed the laws and the state of the world at t . In worlds with deterministic laws, as described in section 1, if an agent is determined at t not to do y at t' , then for any other time t_* ($t_* \leq t$), the agent is determined at t_* not to do y at t' , so the additional time index provides no information. Yet one may conceive of worlds with periods of deterministic evolution interrupted by nondeterministic evolution, in which case the additional time index does provide information. Since such worlds are not our concern here, it is not worth including the additional time index for generality until it is most convenient to do so in section 5.

have (to have) been different relative to any prior t , all the way back to the Big Bang or beyond. As Leibniz ([1686] 1989, 73) wrote, “The whole universe with all its parts ... would have been different from the beginning, if the least thing in it had happened differently than it did.”

We can now state the Conditional Argument in a general form.¹² Simply put, (BT) states that in a deterministic world, a different action requires a different past. But (FP) states that one cannot perform any action that requires a different past, for the past is already fixed, over and done with. The conclusion is that if determinism is true, then one cannot perform any actions other than those one is determined to perform. Determinism is incompatible with the freedom to do otherwise. Incompatibilism is true and compatibilism is false.¹³

Faced with the Conditional Argument as stated above, all compatibilists must reject either (BT) or (FP). The issue of whether to reject (BT) divides the logical space of compatibilists in two. On one side, *miracle compatibilists* reject (BT).¹⁴ According to the standard version of Miracle Compatibilism, if Themistocles had ordered the fleet to Corinth at t' , then all of history would have been the same until shortly before t' , at which time a violation of the actual laws of nature, a so-called divergence miracle, would have allowed Themistocles to order the fleet to Corinth.¹⁵ On the other side are all other compatibilists, *backtracking*

¹² I depart from Fischer’s statement of the Conditional Argument (Fischer 1994, chap. 4) insofar as his does not use the Backtracking Principle (BT) but rather a Principle of the Fixity of the Laws.

¹³ For further discussion of this argument, see Fischer 1994, chap. 4; Ekstrom 2000, chap. 2; and O’Connor 2000, chap. 1.

¹⁴ Lewis (1981) presents the classic defense of what Fischer (1994, 69) calls “Local Miracle Compatibilism,” which I consider the standard version of Miracle Compatibilism. According to my classification, Vihvelin (2000) and Beebe and Mele (2002) are also miracle compatibilists.

¹⁵ Note that if performing an action y at t' requires a divergence miracle *before* t , then according to (FP) the agent cannot at t do y at t' . Put differently, an agent’s “window of opportunity” to perform an

compatibilists, who reject miracles and accept (BT).¹⁶ Since I take (BT) for granted, it is Backtracking Compatibilism that is at stake in what follows. If one establishes (FP), one refutes this version of compatibilism.

3. Backtracking Compatibilism

The principle (FP) introduced above has two components. The first component is the backtracking conditional:

If *s* were to do *y* at *t'*, the past relative to *t* would (have to) be different.

The second component is an associated can-claim:

s cannot at *t* do *y* at *t'*.

The question is whether the truth of the backtracking conditional is compatible with the truth of a *positive* can-claim: *s can* at *t* do *y* at *t'*. According to (FP), they are not compatible. According to Backtracking Compatibilism, they are; even if it is true that if Themistocles had ordered the fleet to Corinth, the past would have (to have) been different all the way back to the Big Bang, Themistocles *could* have done so, provided he was under no coercion or compulsion to issue the order to Salamis.

To be clear, backtracking compatibilists do not claim that Themistocles had an incredible power to *change* the past, to undo events that had already occurred in history. They claim that he had a more modest power, a power to do something like ordering the fleet to Corinth at *t'*, for

 action at *t'* closes at that time *t*, if there is one, such that a divergence miracle would occur before *t* were the agent to perform the action at *t'*. This demonstrates the relevance of (FP) even for miracle compatibilists who deny (BT).

¹⁶ Saunders (1968) presents the classic defense of Backtracking Compatibilism, which Fischer (1994, 79) calls “Multiple-Pasts Compatibilism.” Perry (2004, 2008) also endorses this view. See Fischer 2008 for a reply to Perry.

which the following holds: if he had done it at t' , certain events that actually occurred before t' would never have occurred at all. Different events would have occurred instead. But no event would have both occurred and then been undone by Themistocles.

According to backtracking compatibilists, once we realize that compatibilism does not involve any commitment to a power of undoing the past, (FP) loses any appeal it may have had.¹⁷ What is right about the intuitive idea of the fixity of the past, according to these compatibilists, is that we cannot *undo the past*; what is wrong about (FP) as an expression of this intuitive idea is that (FP) goes further, claiming that we cannot do anything that *requires a different past* either.

Though Fischer (1994, 79) draws a different moral from the distinction between the two types of power over the past, he does mark it by distinguishing between a causal and a noncausal version of (FP). With the modifications as before, the two versions are:

(FPc) For any action y , agent s , and times t and t' ($t \leq t'$), if it is true that if s were to do y at t' , s would thereby initiate a causal sequence issuing in the nonoccurrence of some event e that actually occurred in the past relative to t , then s *cannot* at t do y at t' .

(FP) For any action y , agent s , and times t and t' ($t \leq t'$), if it is true that if s were to do y at t' , the past relative to t would (have to) be different, then s *cannot* at t do y at t' .¹⁸

Let an *action that changes the past* be an action that meets the action type description in (FPc): an action such that if s were to do it at t' , s would thereby initiate a causal sequence issuing in the nonoccurrence of some event e that actually occurred in the past relative to t . Let an *action that*

¹⁷ Foley (1979), Horgan (1985), and Narveson (1977) all defend compatibilism by making this distinction.

¹⁸ Fischer calls this principle (FPnc) for “noncausal.” For simplicity, I call it (FP).

is inconsistent with the past be an action that meets the action type description in (FP): an action such that if s were to do it at t' , the past relative to t would (have to) be different.¹⁹

It is relatively uncontroversial that one cannot perform an action that changes the past. Against Backtracking Compatibilism, Fischer (1994, 79) argues that since one cannot perform an action that changes the past, it is plausible that one cannot perform an action that is inconsistent with the past either. However, this claim is controversial.²⁰ Backtracking compatibilists reject (FP), accepting only the following weaker principle:

(FPw) For any action y , agent s , and times t and t' ($t \leq t'$), if it is true that if s were to do y at t' , the past relative to t would (have to) be different, then s *will not* do y at t' .

Backtracking compatibilists agree that the past is fixed in a strong sense, but they believe that this sense is captured by (FPc) and (FPw), not (FP). They believe that one *can* perform an action that is inconsistent with the past, so (FP) is false, even though one *will not*, so (FPw) is true.

¹⁹ Several points about backward causation are relevant here. First, an action that changes the past involves more than backward causation. Backward causation by an agent would involve the following: an event e occurs before t , an agent performs an action y at a later time t' , and the performance of y at t' is a *cause* of e . Whether or not this is possible in any sense, changing the past would involve even more: an event e occurs before t , an agent performs an action y at t' , and the performance of y at t' causally brings it about that e never occurred. Even if we remove the assumption of causation, there is still a problem: as of a time before the agent's action at t' , the past relative to t was one way, while as of a time after the agent's action at t' , the past relative to t was another way. If the agent's action is a *cause* of the change, as in (FPc), this is a special case of the more general problem. (I thank an anonymous referee for this point.) Finally, note that an action that is inconsistent with the past involves neither backward causation nor the more general problem. Performing an action that is inconsistent with the past would require the past to be different, but it would neither *cause* the past to be different nor involve any difference between how the past relative to t was as of a time before t' and how the past relative to t was as of a time after t' .

²⁰ For arguments that (FPc) does not support (FP), see Kapitan 1996, sec. 4 and Perry 2008, sec. 3.

There are several purported counterexamples to (FP) in the literature.²¹ We can adapt our running example to their common structure. If we ignore the issue of determinism, then intuitively, Themistocles could have ordered the fleet to Corinth. He had the authority to do so, the Peloponnesians wished to fight there anyway, and so on. Now suppose, in a fictional twist, that the only scenario in which Themistocles really would order the fleet to Corinth is one in which Apollo's oracle at Delphi had (contrary to fact) prophesied Persian weakness at Corinth. Let us assume that this is so, not because the oracle can really divine future events, but because Themistocles is so devoted to Apollo that he only goes to battle where the oracle prophesies a Greek advantage. Then we have that

Themistocles can at t order the fleet to Corinth at t' ,

and yet

if Themistocles were to order the fleet to Corinth at t' , then the oracle would have made a different prophecy before t .

According to backtracking compatibilists, there is nothing problematic about the joint truth of the positive can-claim and the backtracking conditional. It is true that the oracle did not prophesy Persian weakness at Corinth and that Themistocles cannot travel back in time to change that. And we are assuming that the only scenario in which Themistocles really would order the fleet to Corinth is one in which the prophecy had been different. So he *will not* order the fleet to Corinth. But what does this show about whether he *can* order the fleet to Corinth?

Backtracking compatibilists say "nothing." All that follows from the assumptions of the case is that Themistocles will not order the fleet to Corinth, as (FPw) implies, not that he cannot,

²¹ Saunders 1968 is a source of such examples. Fischer (1994, 80-83) discusses two examples from Saunders and a third of his own.

as (FP) implies. Against this position, Fischer (1994, 82) argues that in cases such as these, it is not clear that the backtracking conditional is really true. Is it really true that if Themistocles were to order the fleet to Corinth at t' , then the oracle would have made a different prophecy before t ? Fischer's idea is that perhaps the past relative to t would have been the same, while shortly before t' Themistocles would have had a change of character, ignored the oracle for once, and ordered the fleet to Corinth. (Remember that the example does not assume determinism.) One might try to rule out Fischer's possibility by holding that Themistocles could not have overridden his disposition to follow the oracle, but then it is not clear that he *could* have ordered the fleet to Corinth, given the oracle's prophecy about Salamis.

The question is whether we can find a case in which a backtracking conditional is clearly true and in which we nonetheless assent to a positive can-claim. Fischer is skeptical, but he concludes that this question leads to a "Dialectical Stalemate" (83) of the sort encountered in some perennial philosophical debates.

4. The Action-Type Argument for the Principle of the Fixity of the Past

The stalemate can be broken. For it is possible to use (FPw) against Backtracking Compatibilism in a direct argument for (FP). The argument turns on following the type of action introduced above, *action that is inconsistent with the past*, across possible worlds.²²

To fix the intended meaning of 'action type', consider another example. In a dispute about where to send the Greek fleet, the Spartan commander Eurybiades raised his staff, threatening to strike Themistocles. This was an *act*, a particular, unrepeatable doing on the part

²² I share Saul Kripke's reservation about the term 'possible world', given its suggestion of some kind of foreign land (à la Lewis), which the term 'possible history of the world' might avoid (Kripke 1972, 48n15). Nonetheless, I use possible worlds talk for convenience in what follows.

of Eurybiades. It was also a performance of a repeatable *action*. The action was *raising a staff*. Many particular acts, having nothing to do with Eurybiades, have been performances of this action. Moreover, the action *raising a staff* falls under the action type *action involving movement of the arms*. Many actions, having nothing to do with staffs, fall under this type. Hence we have a hierarchy in which every *act* is a spatiotemporally located performance of some action, where an action is distinct from any of its performances—so there can be actions that are not performed—and an action itself may fall under various action types.²³

Many actions fall under the type *action that is inconsistent with the past*. Indeed, if the Backtracking Principle (BT) of section 2 is true, every action that an agent is determined not to perform is an action that is inconsistent with the past. For example, if (BT) is true and Eurybiades is determined by the laws and the past not to strike Themistocles with his staff, then his striking Themistocles counts as an action that is inconsistent with the past. So does the action of Themistocles ordering the fleet to Corinth, if he is determined to order the fleet to Salamis.²⁴

²³ It may help to frame the distinction in terms of *functions*, to which we return in the main text. Corresponding to an action is a function that takes a pair of a possible world and a time and returns a set of acts, which are the performances of the action in that world at that time. Corresponding to an action type is a function that takes a world-time pair and returns a set of actions, or alternatively, a *set of functions from world-time pairs to sets of acts*. One performs an action, but one does not perform an action type. Rather, one performs actions *of* the type. If it can be said that one “performs” an action type at all, it is only *by* performing an action of the type that one does so.

²⁴ Note that one should not speak of an “*act* that is inconsistent with the past.” For acts are occurrences, and no occurrence in a world is inconsistent with the past of that world. What may be inconsistent with the past of the world are various *actions* that are never performed in the world, but are nonetheless contemplated in the course of planning. Assuming (BT) and determinism in the actual world, the action of Themistocles’s ordering the fleet to Corinth, advocated by the Peloponnesians, is inconsistent with the past; however, since Themistocles did not perform this action, there was no *act* of Themistocles’s ordering the fleet to Corinth in the actual world.

Reasoning about action types, not just actions, comes naturally in reasoning about what agents can and cannot do. Suppose that as commander of the Athenian Navy, Themistocles vows, “I cannot do anything that would harm a fellow Greek.” Further suppose that what he says is literally true, without exception. If this is so, then it is natural to reason as follow:

Themistocles cannot do anything that would harm a fellow Greek.

Striking Eurybiades would harm a fellow Greek.

Therefore, Themistocles cannot strike Eurybiades.

Or to put the argument in more artificial language:

Themistocles cannot perform any action whose performance would harm a fellow Greek.

Striking Eurybiades is an action whose performance would harm a fellow Greek.

Therefore, Themistocles cannot perform the action of striking Eurybiades.

Here ‘action whose performance would harm a fellow Greek’ refers to an *action type*, while ‘striking Eurybiades’ refers to an *action* that falls under the action type in the given circumstances. If Themistocles were to perform the action of striking Eurybiades, then that particular episode of violence would be an *act*.

The argument below has a character similar to that of the argument about Themistocles and Eurybiades above. The first step is to establish that an agent cannot perform any action of a certain type, namely *action that is inconsistent with the past*. The next step is to identify an action, such as Themistocles’s ordering the fleet to Corinth, which is of that action type. The conclusion is that Themistocles cannot perform that action. The heart of the argument is in the first step of establishing that one cannot perform any action that is inconsistent with the past.

Since the argument below involves quantification over possible worlds, we must be careful with the action type description ‘action that is inconsistent with the past’. To which past,

actual or possible, does ‘the past’ refer? To be more precise, we need to distinguish two types of action. In general, for each type of anything, there is a corresponding function that takes as an argument a possible world (or a world-time pair) and returns as a value the set of all x such that x falls under the type in that world (at that time).²⁵ The same holds for action types. Since we are interested in actions that are inconsistent with the past *relative to some time*, for each action type we will consider the corresponding function that takes as an argument a world-time pair and returns as a value the set of actions that fall under the action type in that world at that time.

Consider two types of action, F and I , and their corresponding functions, where w is any possible world and $w_{@}$ is our 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 .

For a deterministic world w in which the Backtracking Principle (BT) is true, an action y is in $I(w, t)$ if (and only if) the laws and past relative to t of w determine that y will not be performed in w . (Below we will be explicit about performance times.) In general, y is in $I(w, t)$ if and only if the following backtracking conditional is true *in w* : if y were performed, then the past relative to t would (have to) be different from that of w . This explains F as well, given $F(w, t) = I(w_{@}, t)$.

Assuming determinism and (BT) are actually true, Themistocles’s ordering the fleet to Corinth at t' falls under both action types F and I in the actual world (at any prior time) because $F(w_{@}, t) = I(w_{@}, t)$ = the set of actions inconsistent with the past relative to t of $w_{@}$. However, it is only guaranteed that the equation $F(w, t) = I(w, t)$ will hold in the special case where $w = w_{@}$. For

²⁵ I am neutral on the question of whether types should be *identified* with such functions, but see Egan 2004 for arguments that properties should be identified with functions from world-time pairs to extensions. I am also not committed to the thesis that for *every* such function there is a corresponding type.

arbitrary w , there is no such guarantee. For while the function F returns the same value, no matter which world it takes as an argument (for a given time), the function I may return different values depending on which world it takes as an argument (for a given time).

For example, suppose we ask: If Themistocles had not sent his slave to the Persian camp on the night before the Battle of Salamis, could he have ordered the fleet to Corinth instead of Salamis the next day? We might answer that it depends on whether ordering the fleet to Corinth would have been inconsistent with the past. Here we do not mean inconsistent with the past of the actual world; we mean inconsistent with the past of that possible world in which Themistocles does not send his slave to the Persian camp. In other words, we are saying that the answer depends on whether, in that possible world, Themistocles's ordering the fleet to Corinth at t' falls under action type I . We already know that in that possible world—indeed in *any* possible world—Themistocles's ordering the fleet to Corinth at t' falls under action type F , since in any world that action is an action that is inconsistent with the past of our actual world.

It may turn out that while Themistocles's ordering the fleet to Corinth at t' falls under action type I in the actual world, it does not fall under action type I in the possible world in which Themistocles does not send his slave to the Persian camp. This reflects the fact that whether an action falls under action type I is a contingent matter. This is as it should be since an action that *is* inconsistent with the past *might not have been*, had history gone differently. But whether an action falls under action type F is not a contingent matter in this way. An action that falls under action type F in one world (relative to a time) falls under action type F in all worlds (relative to that time). It is for this reason that I will disambiguate 'action that is inconsistent with the past' by taking it to refer to action type I rather than action type F . To have intuitive descriptions of both types, one could use 'action that is inconsistent with the past of *our* world'

for type *F* and ‘action that is inconsistent with the past of the *agent*’s world’ for type *I*, but it is best to remember the distinction in terms of the corresponding functions given above.

There are many other action types for which it is a contingent matter whether an action falls under the type. For example, consider the action type *action that requires at least n joules of energy* for some specific n . An action such as lifting a large stone on Earth might fall under this action type in the actual world. However, if Earth had been less massive, then lifting the stone might have required fewer than n joules of energy, in which case the action would not have fallen under the type. In general, whenever an action’s falling under a type in a world depends on contingent features of the world, it may then be contingent that the action falls under the type.

In addition to the modal question of whether a proposition of the form *action y falls under type X (relative to time t)* is contingent or necessary, there is another temporal question. Let us adopt John Perry’s distinction between the *truth* of a proposition, which is a timeless property of the proposition, and *its being made true* by events, which is a property of the proposition instantiated at a time (Perry 2004, 234ff). To indicate the timeless property, Perry says that a proposition “be true” or “be false.” To indicate the time-dependent property, he says that a proposition is made true by events up to a time. With this distinction there are three categories of propositions, relative to a time t :

- (i) Propositions that *be true* and that are made true by events up to t .
- (ii) Propositions that *be true* but that are not made true by events up to t .²⁶

²⁶ One might reject the notion that a proposition can *be true*, even though (a) it has not yet been made true but (b) it is the sort of proposition that can be made true. Perhaps those inclined toward indeterminism about the laws or presentism about time will reject propositions of category (ii). Such a view is compatible with everything that follows. Whatever the status of propositions characterized

(iii) Propositions that *be true* but are not the sort of propositions that are made true by events.

Let t be the time of the Battle of Salamis in 480 BCE. Then the proposition that *Xerxes orders his fleet into the straits of Salamis* be true and is made true by events up to t , so it falls under category (i). The proposition that *the reign of Xerxes ends in 465 BCE* also be true, but it is not made true by events up to t . It falls under category (ii) because important events that lead to the end of the reign of Xerxes have not occurred by t . By contrast, the proposition that *the sum of the interior angles of a triangle in Euclidean space equals 180 degrees* be true, but it is not the sort of proposition that is made true by events, so it falls under category (iii). Finally, following Perry (2004, 236), let us introduce the notion of a proposition being *settled* at a time:

A proposition p is settled at time t if and only if p is entailed by propositions of type (i) and (iii).

With this definition, the relevant temporal question is: if it *be true* that *action y falls under type X relative to time t* , when is this *settled*? As explained below, the answer is not necessarily “at t .”

Having introduced some of the modal and temporal distinctions needed for the precise statement of the argument for (FP), let us begin with a rough sketch of the argument: If there is no world in which an agent performs an action of a certain type, then no agent *can* perform an action of that type; and since there is no world in which an agent performs an action that is inconsistent with the past (an action of type I), it follows that no agent can perform an action of that type, which establishes (FP); finally, assuming (BT) and determinism, Themistocles’s ordering the fleet to Corinth is an action that is inconsistent with the past, so he cannot perform it.

by (a) and (b), they are of a different category than (i) or (iii). All that is necessary for our purposes is the three-part distinction.

With the rough sketch in mind, let us start from the beginning of the argument with an imprecise but intuitive version of the first premise:

- (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*.

Strictly we should not talk of “an action of type *X*,” since we need to know at what world-time pairs the action falls under type *X* and when this is settled in a given world. For now, however, we will consider an example in which these subtleties do not matter. According to Greek myth, the hero Perseus was thrown into the ocean as a child, locked in a wooden chest. As the story goes, a fisherman found the wooden chest, opened it, and saved Perseus. But let us change the story. Suppose that Perseus is thrown into the ocean, not in a wooden chest, but in a Hephaestian chest, forged by the god of metallurgy, Hephaestus. Further suppose that in all possible worlds, no agent ever escapes from a Hephaestian chest. Can Perseus escape from a Hephaestian chest? If he can, then there should be some possible world that “witnesses” the truth of this can-claim. Yet by assumption there is no such world, so Perseus cannot escape. This conclusion evidently holds for any sense of the word ‘can’. There is no sense of ‘can’ according to which Perseus can escape from a Hephaestian chest.

Let us deal immediately with an apparent counterexample to (1). Suppose it is impossible for an agent to get into a Hephaestian chest unless he or she has no desire to escape from it. As a consequence, there is no possible world in which an agent escapes from a Hephaestian chest. Nonetheless, several heroes cast into Hephaestian chests have the strength to break out of them and would do so if they desired. Is it not then true that the heroes can escape? If so, we must reject (1), since (1) implies that they cannot.

The objection can be put in a more general form. Suppose there is no possible world in which an agent purposely does something that he or she has no desire, reason, or intention to do. It follows from (1) that an agent cannot do such a thing. Yet it seems that agents often can do what they have no desire, reason, or intention to do. For example, Themistocles may have no desire, reason, or intention to compliment a Spartan general. Still, it may be true that he *can* compliment the general. So even on the assumption that there is no possible world in which an agent purposely does something that he or she has no desire, reason, or intention to do, it seems false that an agent cannot do such a thing. So why should we accept (1), which has this false consequence?

The reply to the objection relies on the following distinction. It is one thing to say that an agent *s* has at *t* no desire, reason, or intention to do *y* at *t'* and yet *s* can at *t* do *y* purposely at *t'*. It is a very different thing to say that *s* can at *t* do *y* purposely at *t'* *while at *t'* having no desire, reason, or intention to do so*. In the second case, it is much more plausible to say that an agent cannot do what is described. And that is the only consequence that the objection manages to draw from (1). For the assumption of the objection must be that there is no possible world in which an agent *s* purposely does something at *t'* for which *s* has at *t'* no desire, reason, or intention to do it. It cannot be that there is no possible world in which an agent has at *t* no desire, reason, or intention to do something at *t'* and yet does that very thing purposely at *t'*; for that happens all the time, simply because the agent develops the right motivation between *t* and *t'*.

Returning to the case of the Hephaestian chest, it is not enough for the purposes of the objection to assume that it is impossible for an agent to get into a Hephaestian chest unless he or she has no desire to escape. After all, someone who initially has no desire to escape might develop such a desire upon realizing how cramped a Hephaestian chest really is. What we have

to assume instead is that it is impossible for an agent in a Hephaestian chest to ever develop a desire to escape. But if that is so, then it is not at all clear that heroes trapped inside can escape, however strong they may be. To take a different example, imagine a siren in whose presence it is impossible for anyone to develop a desire to leave her. Can someone in the presence of such a siren purposely leave her? One cannot, just as one cannot escape from a Hephaestian chest.

Three features are missing in our preliminary statement of premise (1). First, we have not built in the double time indexing of ‘can’, introduced in section 2. Second, we have not specified at what world-time pair the action referred to in (1) falls under the action type X , and third, we have not indicated when this is *settled*. Filling in these gaps, the first premise becomes:

(1*) An agent *cannot* in world w at time t perform an action y at t' ($t \leq t'$) if it is settled in w at t that y falls in w at t under any action type X for which there is no possible world w_* and times t_* and t_*' ($t_* \leq t_*'$) such that an agent performs an action in w_* at t_*' that falls under X in w_* at t_* .

The importance of keeping track of world-time pairs can be seen as follows. Recall that for some action types, such as *action that requires at least n joules of energy*, it is a contingent matter whether an action falls under the type. Suppose one were to claim that for some particular n , there is no possible world in which a (human) agent performs an action that requires at least n joules of energy. Setting aside the plausibility of such a claim, it would follow by (1) that no agent can perform an action that requires at least n joules of energy. Hence if lifting a large boulder on Earth requires at least n joules of energy, then no agent can perform this action. Yet one might reply that if Earth had been less massive, then lifting the boulder would have required fewer than n joules of energy. Suppose there is a possible world v in which Earth is less massive and some agent lifts the boulder on Earth in v . Then since lifting the boulder on Earth in fact

requires at least n joules of energy, and since there is a possible world v in which an agent lifts the boulder on Earth, there is a possible world in which an agent performs an action that requires at least n joules of energy, in which case (1) does not apply. Clearly, this is a fallacy. For the action of lifting the boulder on Earth does not fall under the type *action that requires at least n joules of energy* in the world v , so v is not a counterexample to the claim that there is no possible world in which an agent performs an action of that type. Representing the world-time pairs explicitly in (1*) helps us avoid this fallacy.

Turning to the double time indexing in (1*), consider the action type *action forbidden by the gods*. Suppose that in all possible worlds, no agent ever succeeds in performing an action forbidden by the gods. That is, no agent ever succeeds in performing an action *after* it is forbidden by the gods of his or her world (gods who exist in time and whose prohibitions are permanent, we may imagine), though agents may succeed in performing actions that are only later forbidden. It follows from (1*) that if an action is forbidden by the gods as of now in our world, then no one in our world can now perform that action at any later time.

However, it does not follow from (1*) that if an action is forbidden at time t' , then an agent cannot at t ($t < t'$) perform the action at t' or later. Suppose that at t there is a possible evolution of the world, consistent with the laws of nature and the state of the world at t , in which the agent does something—makes sacrifices or gives offerings—to successfully prevent the gods from forbidding the action by t' . Then it is not *settled* at t that the action falls at t under the action type *action forbidden by the gods at t'* , even if it *be true* that the action is forbidden at t' in the actual evolution of the world. Yet if it is not settled at t , then (1*) does not apply.

The foregoing observations provide a reply to another objection to the informal statement of (1). According to the objection, since no agent performs an action of the type *action that the*

agent will not perform, it follows from (1) that an agent cannot at t perform any action at t' that he or she will not in fact perform at t' . This fatalistic conclusion follows even if we assume that determinism is false, so (1) seems to prove too much, too easily. The reply to this objection is that if determinism is false, then it is not in general settled at t that an action falls at t under the action type *action that the agent will not perform*. Yet if it is not settled at t , then (1*) does not apply. The fatalistic conclusion that the agent cannot at t perform action y at t' , even though it is not yet settled as of t whether y falls under the action type X , is not licensed by (1*). For whether y is an action that the agent will not perform depends on propositions made true *in the future*, which may not be settled at the present time, and (1*) does not commit us to any view about how such propositions constrain the present abilities of an agent, if at all.²⁷

When I state the full argument for (FP) below, it will be helpful to do so in symbols as well as natural language, so let us begin with a symbolic translation of (1*). Let $D(s, y, w, t)$ indicate that agent s *does* action y in world w at time t . Let $C(s, w, t, y, t')$ indicate that agent s *can*, in world w at time t , do action y at time t' . Since we are modeling action types as functions from world-time pairs to sets of actions, let $X(w, t)$ be the set of actions that fall under action type X in world w relative to time t . Hence we will write $y \in X(w, t)$ to indicate that action y falls under action type X in w relative to t . Finally, to indicate that a proposition p is *settled* in world w at time t , we will write $S(p, w, t)$. Hence $S(y \in X(w, t), w, t)$ indicates that it is settled in w at t that $y \in X(w, t)$.

There is one final piece of information that we will encode in our notation. Suppose that if Themistocles were to order the fleet to Corinth at t' , the past relative to t would (have to) be

²⁷ See Perry 2006 for a critique of fatalism using the distinctions from Perry 2004 about when propositions are made true.

different, while if Themistocles were to order the fleet to Corinth at some *other* time t'' , the past relative to t would not (have to) be different. We can think of this in various ways.²⁸ For one, we may think that just as the action's falling under type I is relative to a world w and a time t (as in *the past relative to t*), it is also relative to a candidate performance time t' , in which case we can write $o \in I(w, t, t')$ and $o \notin I(w, t, t'')$, where o stands for the action of Themistocles's ordering the fleet to Corinth. Then to indicate that it is *settled* in w at t that Themistocles's ordering the fleet to Corinth at t' is inconsistent with the past relative to t of w , we write $S(o \in I(w, t, t'), w, t)$.

Given our notation, we can render (1*) symbolically as

$$\forall X [\neg \exists s, w, t, y, t' (t \leq t' \wedge y \in X(w, t, t') \wedge D(s, y, w, t)) \rightarrow \\ \forall s, w, t, y, t' ((t \leq t' \wedge S(y \in X(w, t, t'), w, t)) \rightarrow \neg C(s, w, t, y, t'))],^{29}$$

where the leading quantifier binds an action type variable and the other quantifiers bind agent, world, time, and action variables.³⁰ While the symbolic expression does not exactly mirror the grammatical structure of (1*), it clearly expresses the same principle, which relates what no one *does* in any possible world to what no one *can* do in any world.

²⁸ One possibility is to think of ordering the fleet to Corinth at t' as one action, $o_{t'}$, and ordering the fleet to Corinth at t'' as a different action, $o_{t''}$, writing $o_{t'} \in I(w, t)$ and $o_{t''} \notin I(w, t)$. Another possibility is to think that the action of ordering the fleet to Corinth falls at t under one action type, *action such that if performed at t' , the past would (have to) be different*, but not under a different action type, *action such that if performed at t'' , the past would (have) to be different*, writing $o \in I_{t'}(w, t)$ and $o \notin I_{t''}(w, t)$. Finally, there is the possibility suggested in the main text. It will not matter here which of the three views we adopt, but for the sake of notational clarity, we adopt the notation in the main text.

²⁹ If we assume that when $t' < t$, an agent cannot at the later time t perform any action y at the earlier time t' , translated as $\forall s, w, t, y, t' (t' < t \rightarrow \neg C(s, w, t, y, t'))$, then we can simplify the premises of our argument by dropping the $t \leq t'$ conjuncts. However, we will retain these conjuncts as a reminder of the intended time order.

³⁰ Strings of the same type of quantifier, as in $\forall x \forall y$, are abbreviated, as in $\forall x, y$.

We turn now to the intuitive statement of the second premise of the argument for (FP):

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

It is here that our distinction between action types *F* and *I* above becomes important. The action type referred to in (2) is action type *I*. To see why this matters, suppose it is determined by the laws of nature and the initial conditions of our world that no one will ever insult the oracle at Delphi. It follows from the Backtracking Principle (BT) of section 2 that insulting the oracle at a time t' is inconsistent with the past of our world relative to any prior time t . However, there *are* possible worlds in which some agent insults the oracle. So is this a counterexample to (2)? It is not, for in *those* worlds, insulting the oracle is not an action that is inconsistent with the past (an action of type *I*). It is inconsistent with *our* past (an action of type *F*), but there is no general problem with someone's performing an action that is inconsistent with the past of *someone else's* world. The problem comes with someone's performing an action that is inconsistent with the past of his or her own world. (Compare the case of *n joules of energy* discussed above.)

Once again, we need to add the missing world and time variables to the premise:

(2*) There is no possible world w in which an agent performs an action at t' that falls in w at t ($t \leq t'$) under the action type *action that is inconsistent with the past* (action type *I*).

As stated, premise (2*) is a consequence of a principle from section 3 that compatibilists accept:

(FP w) For any action y , agent s , and times t and t' ($t \leq t'$), if it is true that if s were to do y at t' , then the past relative to t would (have to) be different, then s *will not* do y at t' .

We read (FP w) as saying that if an action falls under action type *I* in a world, then an agent *will not* perform the action in that world. If we wanted to apply (FP w) only to the actual world, we

could read it as saying something about actions of type F only, but we cannot do so if we want to apply (FPw) to other worlds as well.

For example, suppose it is true that if Themistocles had not sent his slave to the Persian camp on the night before the Battle of Salamis, then if he had ordered the fleet to Salamis the next day, the past leading up to his order would have (to have) been different. In other words, let u be the possible world in which Themistocles does not send his slave to the Persian camp, and suppose it is true in u that if Themistocles were to order the fleet to Salamis, the past leading up to his order would (have to) be different. (Here we obviously do not mean different from the past of $w_{@}$, but different from the past of u .) Then it should follow from (FPw) that Themistocles will not order the fleet to Salamis in u . Yet Themistocles's ordering the fleet to Salamis does not fall under action type F in u , or in any other world for that matter, since it is not inconsistent with the past of our actual world. Rather, Themistocles's ordering the fleet to Salamis falls under action type I in u , for we are assuming that it is inconsistent with the past of u . The desired conclusion, that Themistocles will not perform the action in u , follows from (FPw) when the principle is understood in terms of action type I , not F .

Observe that the symbolic translation of (FPw), so understood, is

$$\forall s, w, t, y, t' ((t \leq t' \wedge y \in I(w, t, t')) \rightarrow \neg D(s, y, w, t')),$$

and the symbolic translation of (2*) is

$$\neg \exists s, w, t, y, t' (t \leq t' \wedge y \in I(w, t, t') \wedge D(s, y, w, t')),$$

which are equivalent. Hence (2*) is a consequence of (FPw). Just as no one ever performs an action that changes the past of his or her world, no one ever performs an action that is inconsistent with the past of his or her world either. And this is all we need to conclude by (1*) that no one *can* perform an action that is inconsistent with the past of his or her world:

$$\forall s, w, t, y, t' ((t \leq t' \wedge S(y \in I(w, t, t'), w, t)) \rightarrow \neg C(s, w, t, y, t')),$$

which is precisely our translation of the Principle of the Fixity of the Past (FP). As advertised at the beginning of this section, we have used (FPw) against Backtracking Compatibilism in a direct argument for (FP). For (FPw) is equivalent to (2*), and (FP) follows from (1*) and (2*).

To prepare for the application of (FP) to our running example, let h stand for Themistocles and o stand for the action of Themistocles's ordering the fleet to Corinth, as before. If it is determined by the laws of nature and the initial conditions of the world that Themistocles will not order the fleet to Corinth at t' , then by the Backtracking Principle (BT), this action is inconsistent with the past relative to any prior t :

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow o \in I(w_{@}, t, t').$$

Moreover, if it is *determined* that Themistocles will not order the fleet to Corinth at t' , then it is *settled* at t that this action is inconsistent with the past relative to t :

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow S(o \in I(w_{@}, t, t'), w_{@}, t).$$

This conditional is the final piece of our picture, which I present in full below. Those who adopt a Humean view of the laws of nature, as miracle compatibilists typically do, may dispute the step above from *determined* to *settled*, since they may deny that the deterministic laws of nature are settled at t . I defer discussion of this dispute to section 5, where I make explicit the view of laws assumed in the arguments here against Backtracking Compatibilism.

We are now ready to bring everything together, to use the Action-Type Argument for (FP) to answer our original question: assuming determinism, could Themistocles have ordered the fleet to Corinth? Below, the natural language premises of the argument are the imprecise versions, while the symbolic expressions include all of the necessary details:

(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 .

$$\forall X [\neg \exists s, w, t, y, t' (t \leq t' \wedge y \in X(w, t, t') \wedge D(s, y, w, t')) \rightarrow \\ \forall s, w, t, y, t' ((t \leq t' \wedge S(y \in X(w, t, t'), w, t)) \rightarrow \neg C(s, w, t, y, t'))]$$

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

$$\neg \exists s, w, t, y, t' (t \leq t' \wedge y \in I(w, t, t') \wedge D(s, y, w, t'))$$

(3) Therefore, an agent cannot perform an action that is inconsistent with the past (an action of type I) (from (1) and (2)).

$$\forall s, w, t, y, t' ((t \leq t' \wedge S(y \in I(w, t, t'), w, t)) \rightarrow \neg C(s, w, t, y, t'))$$

Having established the fixity of the past in (3), we continue to the case of Themistocles:

(4) If determinism is true, then Themistocles's ordering the fleet to Corinth at t' is an action that is inconsistent with the past (an action of type I) (application of (BT)).

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow S(o \in I(w_{@}, t, t'), w_{@}, t)$$

(5) Therefore, if determinism is true, Themistocles cannot order the fleet to Corinth at t' (from (3) and (4)).

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow \neg C(h, w_{@}, t, o, t')$$

The argument generalizes for any action that one does not perform in life, with the conclusion that if determinism is true, one cannot do anything in life other than what one ends up doing. The argument is valid. If it is also sound, as I have argued, then it refutes Backtracking Compatibilism.

The classic compatibilist move is to point out that Themistocles did not order the fleet to Corinth in the actual world *because he did not want to*. He would have *if* he had wanted to.

Hence there is an appropriate possible world in which he wants to order the fleet to Corinth and *does so*, thereby performing an action that is inconsistent with the past of our world. So far, so good. The problem comes when compatibilists assume that since there is a possible world in which an agent performs an action that is inconsistent with the past in the sense of action type F , Themistocles can in the actual world perform such an action and order the fleet to Corinth at t' . They have neglected the fact that in the actual world, Themistocles's ordering the fleet to Corinth at t' is also an action that is inconsistent with the past in the sense of action type I , and there is no possible world in which an agent performs an action of type I , so Themistocles *cannot* order the fleet to Corinth at t' . That is the lesson of the Action-Type Argument for the Principle of the Fixity of the Past.

5. A Simple Argument for Incompatibilism

In this section, we use the structure of the Action-Type Argument for (FP) to give a still more direct argument for incompatibilism, which does not go through (FP) and does not assume the Backtracking Principle (BT).

Let DN be the action type *action that the agent is determined not to perform*, so that $DN(w, t, t')$ is the set of actions y such that the laws of world w together with the state of w at t determine that the agent will not perform y at t' . Consider the following Simple Argument for incompatibilism:

- (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 .

$$\forall X [\neg \exists s, w, t, y, t' (t \leq t' \wedge y \in X(w, t, t') \wedge D(s, y, w, t')) \rightarrow$$

$$\forall s, w, t, y, t' ((t \leq t' \wedge S(y \in X(w, t, t'), w, t)) \rightarrow \neg C(s, w, t, y, t'))]$$

(2b) There is no possible world in which an agent performs an action of the type *action that the agent is determined not to perform* (from the definition of ‘determined’).

$$\neg \exists s, w, t, y, t' (t \leq t' \wedge y \in DN(w, t, t') \wedge D(s, y, w, t'))$$

(3b) Therefore, an agent cannot perform an action of the type *action that the agent is determined not to perform* (from (1) and (2b)).

$$\forall s, w, t, y, t' ((t \leq t' \wedge S(y \in DN(w, t, t'), w, t)) \rightarrow \neg C(s, w, t, y, t'))$$

(4b) If determinism is true, then Themistocles’s ordering the fleet to Corinth at t' is an action of the type *action that the agent is determined not to perform*.

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow S(o \in DN(w_{@}, t, t'), w_{@}, t)$$

(5b) Therefore, if determinism is true, Themistocles cannot order the fleet to Corinth at t' (from (3b) and (4b)).

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow \neg C(h, w_{@}, t, o, t')$$

The argument is valid. I have argued that premise (1) in its precise form is true. Given that s is *determined not to do y* entails s will not do y, premise (2b) is also true. Whereas premise (2) of the argument for (FP) required (BT), neither (2b) nor the other premises of the Simple Argument require (BT). Rather, what the argument requires is a particular view of the laws of nature.

On a “governing” view of the laws of nature, the propositions expressed by statements of laws fall into Perry’s category (iii): they be true, but they are not *made true* by events. A law of nature is something that governs events as they unfold, so events fall into the patterns they do because the laws are what they are. On a Humean view of the laws, the propositions expressed by statements of laws fall into Perry’s category (ii): they are made true by events, only they are not made true until the cascade of events is complete, if it ever is. Here the relationship between

laws and events is reversed; a law of nature is just a regularity that turns out to be exceptionless over all time, so the laws are what they are because events falls into the patterns they do.³¹

It is a consequence of the Humean view of laws that even if it *be true* that Themistocles is determined by the laws of nature and the initial conditions of the world at t_0 not to order the fleet to Corinth at t' , it may not be *settled* at t ($t_0 \leq t < t'$) that he is so determined. The reason is that type (ii) propositions expressing the laws themselves are not made true until the cascade of events is complete, long after t . Given this view of laws, the action type *action that the agent is determined not to perform* is similar to the fatalistic action type *action that the agent will not perform*, mentioned in section 4. For it may not be settled at t whether an action falls under either of these types, since this depends on what happens in the *future* relative to t —on what the laws will turn out to be, on what the agent will turn out not to do—not just on what happened in the past relative to t . In other words, while no one can deny the (definitional) conditional

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow o \in DN(w_{@}, t, t'),$$

a Humean about laws of nature might deny the conditional

$$(t \leq t' \wedge \text{determinism} \wedge \neg D(h, o, w_{@}, t')) \rightarrow S(o \in DN(w_{@}, t, t'), w_{@}, t).$$

By contrast, the latter conditional is unproblematic given a governing view of laws. Since propositions expressing the state of the world in the past relative to t are of category (i), if propositions expressing the laws are of category (iii), then it is settled at t that the past and the laws determine that Themistocles will not order the fleet to Corinth at t' .

I will not argue here against the Humean view of laws, typically favored by miracle compatibilists.³² The argument for (FP) and the Simple Argument are to be understood under the

³¹ While it has become standard to call such a view of laws “Humean” (see Beebe and Mele 2002), some scholars have argued that it is a mistake to attribute such a view to Hume. See, for example, Strawson 1989.

assumption of the governing view of laws, which is in line with the backtracking compatibilists' principle (BT). Something like the idea that laws have a modal force beyond that of mere regularities must be behind the idea of (BT), the idea that if history had gone differently, it would not have been because of any miracles, but rather because the past would have been different all the way back. Our conclusion is therefore qualified, but still strong: incompatibilism follows on the assumption of a governing view of laws.

6. Conclusion

To find out whether a person could have done something, we cannot ignore the laws of nature. The arguments of this article prompt two questions: Is it *true* that deterministic laws govern human action? If so, how *bad* is it to lack the freedom to do otherwise?³³

I have addressed only the question of the compatibility of freedom and determinism. In section 4, I confronted Backtracking Compatibilism with a new Action-Type Argument for the Principle of the Fixity of the Past (FP). The objective was to break the Dialectical Stalemate, discussed in section 3, concerning the Conditional Argument for incompatibilism introduced in section 2. In section 5, I used the structure of the Action-Type Argument for (FP) to give a Simple Argument for incompatibilism that bypasses the Conditional Argument. The conclusion was that incompatibilism follows on the assumption of a governing view of the laws of nature.

All of this brings us back to our question from the beginning: assuming determinism, could Themistocles have ordered the fleet to Corinth instead of Salamis? According to the

³² See, for example, Lewis 1973, sec. 3.3 in connection with the Local Miracle Compatibilism of Lewis 1981. See Beebe and Mele 2002 and Perry 2004 for further discussion of how compatibilists might take advantage of a Humean view of laws.

³³ Or as Dennett (1984) asks: Is the freedom to do otherwise “worth wanting”?

arguments of this article, he could not have. If it was determined from the time of the Big Bang that Themistocles would receive the oracular prophecy that he did, that he would quarrel with the Peloponnesians as he did, and that he would order the fleet to Salamis as he did, then he was not free to do otherwise. He was not free to do anything but lead the Greeks to victory, and Xerxes was not free to do anything but lead the Persians to defeat.

There is only one option left for the compatibilist: adopt a Humean view of laws. While it may be consistent to pair such a view of laws with Backtracking Compatibilism, the more natural fit is with Miracle Compatibilism. According to the latter theory, if determinism is true, it would have taken a miracle for Themistocles to send the fleet to Corinth, and yet Themistocles could have sent the fleet to Corinth all the same. Perhaps it will take a miracle to save compatibilism. For those who doubt that one can perform an action, the performance of which requires a miracle, incompatibilism awaits. If deterministic laws govern human action, then you do not have the power to choose, among a number of actions you can perform, the one you will. The only action you can perform is the one you will perform.

For those who regard the traditional debate about the freedom to do otherwise as resolved in favor of incompatibilism, it is time to consider the other kinds of control that we wield over the natural world, including ourselves as parts of it. We may yet find a form of inner freedom that is possible even in a deterministic world.

References

- Beebe, Helen. 2002. "Reply to Huemer on the Consequence Argument." *Philosophical Review* 111, no. 2: 235-41.
- Beebe, Helen and Alfred Mele. 2002. "Humean Compatibilism." *Mind* 111, no. 442: 201-23.

- Bennett, Jonathan. 1984. "Counterfactuals and Temporal Direction." *Philosophical Review* 93, no. 1: 57-91.
- Dennett, Daniel C. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
- Egan, Andy. 2004. "Second-Order Predication and the Metaphysics of Properties." *Australasian Journal of Philosophy* 82, no. 1: 48-66.
- Ekstrom, Laura Waddell. 2000. *Free Will: A Philosophical Study*. Boulder, CO: Westview Press.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will*. Malden, MA: Blackwell.
- Fischer, John Martin. 2008. "My Way and Life's Highway: Replies to Steward, Smilansky, and Perry." *The Journal of Ethics* 12, no. 2: 167-89.
- Foley, Richard. 1979. "Compatibilism and Control over the Past." *Analysis* 39, no. 2: 70-74.
- Ginet, Carl. 1966. "Might We Have No Choice?" In *Freedom and Determinism*, ed. Keith Lehrer, 87-104. New York: Random House.
- Horgan, Terence. 1985. "Compatibilism and the Consequence Argument." *Philosophical Studies* 47, no. 3: 339-56.
- Hume, David. [1777] 1993. *An Enquiry Concerning Human Understanding*, ed. Eric Steinberg. Indianapolis: Hackett.
- Kane, Robert. 2002. "Introduction: The Contours of Contemporary Free Will Debates." In *The Oxford Handbook of Free Will*, ed. Robert Kane, 3-41. New York: Oxford University Press.
- Kapitan, Tomis. 1996. "Modal Principles in the Metaphysics of Free Will." In *Philosophical Perspectives* 10, *Metaphysics*, ed. James E. Tomberlin, 419-45. Malden, MA: Blackwell.
- Kripke, Saul A. 1972. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

- Lamb, James W. 1977. "On a Proof of Incompatibilism." *Philosophical Review* 86, no. 1: 20-35.
- Leibniz, G.W. [1686] 1989. "Remarks on Arnauld's Letter." In *Philosophical Essays*, ed. And trans. Roger Ariew and Daniel Garber, 69-77. Indianapolis: Hackett.
- Lewis, David. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis, David. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13, no. 4: 455-76.
- Lewis, David. 1981. "Are We Free to Break the Laws?" *Theoria* 47, no. 3: 113-21.
- Narveson, Jan. 1977. "Compatibilism Defended." *Philosophical Studies* 32, no. 1: 83-87.
- O'Connor, Timothy. 2000. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
- Peacocke, Christopher. 1999. *Being Known*. Oxford: Oxford University Press.
- Perry, John. 2004. "Compatibilist Options." In *Freedom and Determinism*, ed. Joseph Keim Campbell, Michael O'Rourke, and David Shier, 231-54. Cambridge, MA: MIT Press.
- Perry, John. 2006. "How Real Are Future Events?" In *Time and History, Proceedings of the Twenty-Eighth International Wittgenstein Symposium*, ed. Friedrich Stadler and Michael Stöltzner, 13-30. Frankfurt: Ontos Verlag.
- Perry, John. 2008. "Can't We All Just Be Compatibilists?: A Critical Study of John Martin Fischer's *My Way*." *Journal of Ethics* 12, no. 2: 157-66.
- Saunders, John Turk. 1968. "The Temptations of 'Powerlessness'." *American Philosophical Quarterly* 5, no. 2: 100-108.
- Strauss, Barry. 2004. *The Battle of Salamis: The Naval Encounter That Saved Greece—and Western Civilization*. New York: Simon and Schuster.
- Strawson, Galen. 1989. *The Secret Connexion: Causation, Realism, and David Hume*. Oxford: Oxford University Press.

van Inwagen, Peter. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27, no. 3: 185-99.

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

Vihvelin, Kadri. 2000. "Libertarian Compatibilism." In *Philosophical Perspectives* 14, *Action and Freedom*, ed. James E. Tomberlin, 139-66. Malden, MA: Blackwell.

Wiggins, David. 1973. "Towards a Reasonable Libertarianism." In *Essays on Freedom and Action*, ed. Ted Honderich, 31-62. London: Routledge and Kegan Paul.