

# Freedom and Modality<sup>1</sup>

Wesley H. Holliday

University of California, Berkeley

## 1. ‘Can’ and ‘Possible’

What is the relation between freedom and modality? There is a long history of attempts to answer this question, from philosophers of the Hellenistic period to early modernity to the present. In recent history, Peter van Inwagen’s *An Essay on Free Will* (1983) was a milestone contribution to the question. By presenting a modal version of his now famous Consequence Argument, van Inwagen helped bring the modern possible-worlds picture of modality to the fore in debates about the compatibility of freedom and determinism.

What is the relation between what one *can* do in the actual world and what one *does* in other possible worlds? In “Freedom and the Fixity of the Past” (Holliday 2012, hereafter ‘FFP’), I argued that there is at least (roughly) the following relation:

- (1) For any action type  $X$ , an agent *cannot* perform an action of type  $X$  if there is no possible world in which an agent performs an action of type  $X$ .<sup>2</sup>

In this paper, I will provide some further motivation for (1) and reply to objections to it.

---

<sup>1</sup> Preprint of April 2013. Forthcoming in *Being, Freedom, and Method: Themes from van Inwagen*, ed. John Keller, Oxford University Press.

<sup>2</sup> To be more precise: 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  $t_*$ . For the definition of ‘settled’, a notion due to John Perry, see FFP, 193.

Reasoning from what no one does across possible worlds to what no one can do in a particular world seems natural. Consider an epistemological example: in a world where no one ever knows the complete description *C* of the world, is it nonetheless the case that an agent *can know* or *could have known C*? Chalmers (2012, 49) reasons as follows:

*C*...will have to be sufficiently encompassing that its truths are jointly true of this world and this world alone. But then, assuming that no one actually knows *C*, it will be impossible to know *C*. Any world in which someone knows *C* will differ from the actual world and will therefore be a world in which *C* is false. But there are no worlds in which someone knows *C* and *C* is false. So no one can know *C*.

In the last two sentences, Chalmers makes the natural reasoning step noted above.<sup>3</sup>

What distinguishes (1) from other principles relating freedom and modality is that it is stated in terms of an *action type* (see FFP, 189f). To see how this matters, consider another example from FFP: imagine a Siren so alluring that it is impossible for anyone in her presence to move away from her; as a result, there is no possible world in which someone who hears her song ever escapes her. *Can* someone who hears the Siren's song perform an action such as, e.g., turning around and walking away? I don't think so. Can he *run* away instead? No. Changing the action doesn't help, as long as it is an action that moves the agent from within proximity of the Siren to a place away from her. For by hypothesis there is no world in which an agent performs such an action.

---

<sup>3</sup> Spencer (2013) argues that an agent can know the likes of *C* and can know Fitch-paradoxical propositions of the form '*p* and I do not know that *p*', which may be in *C*, so he thinks there are counterexamples to a principle in the spirit of (1). I side with those like Chalmers who think that *C* and Fitch-paradoxical propositions cannot be known (see, e.g., Williamson 2002, Ch. 12), but I cannot do justice to Spencer's argument here.

To relate this reasoning to (1), let  $M$  be the action type *action that moves the agent from within proximity of the Siren to a place away from her*. As an instance of (1), we have:

(1 $M$ ) If there is no possible world in which an agent performs an action of type  $M$ , then no agent can perform an action of type  $M$ .

Many actions fall under the action type  $M$ , including walking, running, skipping, and jumping away from the Siren. By (1 $M$ ) and our hypothesis, it follows that no agent can walk, run, skip, or jump away from the Siren, in agreement with the previous reasoning. Thus, the generality obtained by stating (1) in terms of an action type allows us to draw a general conclusion about all actions falling under the type.

The fact that we can see a natural pattern of reasoning like that above as an instance of the general form of reasoning according to (1) provides defeasible motivation for (1). Of course, there are less general principles than (1) that we can see the above reasoning as following, e.g., (1 $M$ ) itself, but there are a range of examples accounted for by (1) but not (1 $M$ ), including the two examples to be given below (*insulting the gods* and *lifting the boulder*). It is important that these examples, which can be multiplied, provide motivation for (1) that is independent of any views about the relation between freedom and determinism. I will return to this point below.

It is also important to preempt a general form of objection to (1). Consider another mythological example: suppose there is no possible world in which a mortal performs an action *deemed insulting by the gods* in that world. Further suppose that in the

actual world, the gods deem *dressing up as Zeus* an insult.<sup>4</sup> Two philosophers, **A** and **B**, debate whether a mortal in the actual world can dress up as Zeus:

**A:** Given our assumptions, it is evident that no mortal can do so.

**B:** Please explain.

**A:** Since there is no possible world in which a mortal performs an action deemed insulting by the gods in that world, a mortal *cannot* perform such an action; and since dressing up as Zeus is such an action in our world, none of us can do it.

**B:** Not so fast. I agree that if a mortal can dress up as Zeus, then there is some world in which a mortal *does* dress up at Zeus. But there *are* such worlds: worlds in which the history of relations between gods and mortals is different, so dressing up as Zeus is considered a *compliment* to the gods rather than an insult! You agree that there are such worlds, do you not?

**A:** I do. But surely a world in which a mortal *compliments* the gods by dressing up as Zeus does not serve as a witness for the claim that in our world, where the gods deem dressing up as Zeus an *insult*, a mortal can do so.

**B:** Why not?

**A:** According to your reasoning, in order to show that a mortal can perform an action, such as dressing up as Zeus, which in our world is of some type, such as *action deemed insulting by the gods*, we are allowed to find witnessing worlds in which the action falls under a *different* type, such as *action deemed complimentary by the gods*—even though the very reason for worrying that we

---

<sup>4</sup> We may even suppose that the gods deem *choosing* to dress up as Zeus an insult, so in the following arguments, one may take the action in question to be *choosing* to dress up as Zeus, rather than dressing up as Zeus.

cannot dress up as Zeus was that the gods deem it an insult in our world. My friend, you have committed the famous Fallacy of Switching Types! (Well, it isn't famous yet. But in FFP I suggested that the Fallacy of Switching Types, although not by that name, is one of the mistakes made by philosophers who hold that the *freedom to do otherwise* is compatible with determinism. But more on that shortly.)

**B:** You accuse me of fallacious reasoning. But we agree that there is no possible world where a mortal performs an action deemed insulting by the gods there.

**A:** We do.

**B:** It follows, according to a possible-worlds analysis of counterfactuals (Lewis 1973), that the following counterfactual is true in *our world*, where the gods deem dressing up as Zeus an insult: if a mortal *were* to dress up as Zeus, then the gods *would not* deem it an insult!<sup>5</sup> Then since there *is* a world in which a mortal dresses up as Zeus and it is not an insult, I see no reason yet why we cannot do so.

**A:** I am not sure about your counterfactual. But you are assuming that the only relevant action types are the types that the action *would* fall under if it were performed—the types that it falls under in worlds whose “agential accessibility” from the actual world is precisely what is under dispute—while the action type the action *does* fall under is irrelevant. But this assumption will lead you astray.

**B:** How so?

---

<sup>5</sup> Since by assumption there are no worlds where an agent performs an action deemed insulting by the gods in that world, all of the *closest worlds* where a mortal dresses up as Zeus are worlds where doing so is not an insult.

A: Suppose that in the actual world, lifting a particular boulder on Earth requires at least  $n$  joules of energy.<sup>6</sup> Also assume that there is no possible world in which a human performs an action that requires at least  $n$  joules of energy. Then surely a human in our world cannot lift the boulder. But you reason as follows: by our assumption and the possible-worlds analysis of counterfactuals, if someone were to lift the boulder, the action of lifting the boulder would require *fewer* than  $n$  joules of energy; so when we ask ourselves whether a human can lift the boulder, which in the actual world requires at least  $n$  joules of energy, we should ask ourselves whether there is a possible world in which a human performs an action that requires *fewer* than  $n$  joules of energy! But that is absurd.

In FFP, I argued that some of the problems that **A** has identified with **B**'s reasoning are analogous to problems with compatibilist reasoning in the debate over freedom and determinism (see FFP, 202).

The main argument of FFP shows that a precise version of (1) (see footnote 1 above) entails<sup>7</sup> the incompatibilist view that in a *deterministic* world, agents *cannot* do anything except what they *will* do.<sup>8</sup> I do not have room here to rehearse the argument, so

---

<sup>6</sup> This example is from section 4 of FFP.

<sup>7</sup> Assuming a non-Humean view of the laws of nature. See section 5 of FFP.

<sup>8</sup> Since the precise version only applies to actions and action types at times at which it is *settled* (see FFP, 193) that the action will fall under the action type—for otherwise the agent may do something so that the action will not fall under the type—the principle does not have the fatalistic consequence that even in an indeterministic world, agents cannot do anything except what they will do (see FFP, 197). But here is another consequence of the principle: since there is no possible world in which a backward time traveler performs an action that permanently prevents his own creation, if it is settled at a time that killing

I refer the reader to FFP. As far as I know, no one has disputed the validity of this argument for incompatibilism, but only the truth of (1). In my view, the most promising form of compatibilist response to the argument is the following: propose a modified version of (1) that restricts the quantification over action types in (1) to a (perhaps context-dependent [cf. Horgan 1979]) subclass of action types, identified in a principled way, such that the alternative version explains the initial appeal of (1) and yet avoids the incompatibilist consequence.<sup>9</sup> Another serious response, which applies as much to

---

one of his forebears falls under that action type, then the backward time traveller cannot do so at that time. Some have claimed to the contrary that a backward time traveler *can* do so, but I do not have room to discuss the issue here (see Vihvelin 1995, references therein, and the subsequent literature).

<sup>9</sup> In the dialogue between **A** and **B**, we saw a problem with the following modification: only quantify over the action types that the action would fall under if it were performed. Another problem is that this restriction makes the principle *trivial*: for if  $X$  is an action type that the action would fall under if it were performed (and the action itself is not impossible), then of course (according to possible-worlds analyses of counterfactuals) there is a possible world where an agent performs an action of type  $X$ ; so we would only quantify over action types to which the antecedent of the conditional in the principle does not apply. Here is another proposal that I find problematic: only quantify over the action types that the action would fall under if the agent were to *choose* to perform the action. In its favor, this proposal does not commit the compatibilist to the problem pointed out by **A** in the case of the boulder; for if the agent were to choose to lift the boulder, then the action of lifting the boulder would still fall under the type *action requiring at least  $n$  joules of energy*. Moreover, if this proposal were plausible, it would appear to block the incompatibilist argument of FFP; for in the running example of FFP, if Themistocles were to choose to send the fleet to Corinth, then the action of *sending the fleet to Corinth* might not fall under what I called action type  $I$  (see FFP, 191). However, the argument of FFP can be reformulated in terms of the action of *choosing to send the fleet to Corinth*,

compatibilist principles as to (1), is to argue that there is no objective fact of the matter as to which position is correct, the (contextualist) compatibilist or (invariantist) incompatibilist, due to what Unger (1984, 54ff) calls the *semantic relativity* of ‘can’. However, in what follows I will consider two objections to (1) along different lines.

## 2. Objections

In “Incompatibilism and the Past” (this volume), Tognazzini and Fischer (hereafter ‘T&F’) raise two worries about (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’s impossible, and hence is unmotivated” (cite). I will address these worries in reverse.

### 2.1. “No Independent Motivation”

FFP does not claim that (1) is supposed to capture the intuitive idea that no one can do what is impossible, so I will take T&F’s second worry to be that FFP does not provide, as they put it (cite), “independent motivation” for (1). According to T&F, the problem with (1) is that “it simply crystallizes the incompatibilist intuition” (cite). They seem to mean

---

with the conclusion that Themistocles cannot perform this action when (it is settled that) it is of type *I* (indeed, T&F take *choosing Corinth* to be the action in the argument). The compatibilist might reply that if Themistocles were to *choose* to perform the action of *choosing to send the fleet to Corinth*, then the action of *choosing to send the fleet to Corinth* might not fall under type *I*, so type *I* is not in the relevant subclass of action types. But this appeal to “choosing to choose” makes little sense and leads to a regress.



‘simply’ in the sense of ‘merely’. But I would rather put the point a different way: it crystallizes the incompatibilist intuition in a simple way, which is a virtue. However, I do not agree that the only motivation for (1) comes from incompatibilism itself. To the contrary, section 1 of this paper provides motivation for (1) that is independent of the debate about freedom and determinism.<sup>10</sup> The motivation is of course defeasible, but (1) is not unmotivated.

## 2.2. “Begging the Question”

T&F do not explain what it means for the compatibilist to have “antecedent reason” to reject (1). However, they correctly point out that (1) is inconsistent with standard compatibilist possible-worlds analyses of ‘can’. Moreover, they say that the compatibilist “will deny” it, “will be puzzled” by it, “will not be bothered by it,” and “will maintain that it stacks the deck against him” (cite), because the necessary condition for ‘can’ in (1) is not a necessary condition according to compatibilist analyses.<sup>11</sup> What A called in section 1 the Fallacy of Switching Types, the compatibilist embraces: “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” (T&F, cite). In short, compatibilists are

---

<sup>10</sup> Cf. van Inwagen (1983) on his principle ( $\beta$ ): “the examples I gave in support of ( $\beta$ ) [see 98] did not presuppose the *incompatibility* of free will and determinism” (102).

<sup>11</sup> Cf. van Inwagen (1983) on his principle ( $\beta$ ): “there are almost certainly philosophers who would say that the fact that determinism and compatibilism together entail that there are counterexamples to ( $\beta$ ) shows that my use of ( $\beta$ ) in an argument for incompatibilism is question-begging. But if this accusation were right, it’s hard to see how any argument could avoid begging the question” (102).

happy with their own analysis,<sup>12</sup> in terms of which they can explain why they reject (1). Does this give them an “antecedent reason” to reject (1)? In the sense that they had their analysis *before* they saw (1), it is antecedent. Do they have an “antecedent reason” in a deeper sense?

The beauty of the charge that a premise begs the question is that it promises to provide a way of undermining a valid argument without arguing that any of the premises are false. To show that compatibilists have a *good* reason to reject (1), one would presumably need to show that the analysis of ‘can’ that compatibilists use to reject (1) is a *good* analysis, or give a good counterexample to (1) motivated independently of compatibilism. But to show that compatibilists have an “antecedent reason” to reject (1), this hard work is apparently not necessary. As T&F note (cite), the simple conditional analysis of ‘can’ that they consider is problematic for well-known reasons. But an argument can still beg the question against a bad theory.<sup>13</sup> If it does, is that a bad thing?

It seems to me that we need a different way of thinking about the debate. In his

---

<sup>12</sup> I am reminded of Nozick’s (1981) remark: “Though philosophy is carried out as a coercive activity, the penalty philosophers wield is, after all, rather weak. If the other person is willing to bear the label of ‘irrational’ or ‘having the worse arguments,’ he can skip away happily maintaining his previous belief...” (4).

<sup>13</sup> Does the argument from FFP and this paper beg the question against the compatibilist? According to Fischer and Pendergraft (2013), “an argument begs the question just in case the proponent of the argument has no reason to accept the relevant premise, apart from a prior acceptance of the conclusion” where ‘reason’ does not mean *decisive* reason, but rather a “reason that at least renders it plausible to accept the relevant premise (apart from accepting the conclusion)” (584). But section 1 of this paper at least renders it plausible to accept (1) apart from a prior acceptance of incompatibilism.

fascinating discussion of “Philosophical Failure,” van Inwagen (2006) suggests that we should not think of a philosophical debate as an exchange between two philosophers with opposing views, trying to convert each other. Instead, we should think of the debate on the forensic model: the philosophers’ “purpose is not to convert *each other* but rather to convert the audience—an audience whose members (in theory) bear no initial allegiance to either position, although they regard the question ‘Which of these two positions is correct?’ as an interesting and important one” (44). Viewing a philosophical debate in this way has consequences for the idea of begging the question:

[O]n the model of debate I have endorsed, Norma the nominalist need not worry about whether Ronald the realist will accept her premises. She is perfectly free to employ premises she *knows* Ronald will reject; her only concern is whether the audience of agnostics will accept these premises. Suppose, for example, that she uses the premise, ‘We can have knowledge only of things that have the power to affect us.’ It may well be that no realist, certainly no realist who had thought the matter through, would accept that premise. If Norma tried to use this premise in a debate of the first sort, in an attempt to convert Ronald the realist to nominalism, Ronald would very likely say, ‘Well, of course I don’t accept *that*; that just begs the question against my position.’ But in a debate conceived on the forensic model, Ronald can’t make that response, for the simple reason that what *he* thinks is quite irrelevant to the logic of the debate. If Ronald thinks that there is any danger of the agnostics accepting this premise, it will do him no good to tell the audience that of course no realist would accept this principle and that it therefore begs the question against realism. He’ll have to get down to the business of

convincing the agnostics that they should reject, or at least not accept, this premise. (46)

I think the same points apply to T&F's objection. Rather than claiming that (1) begs the question against the compatibilist, the compatibilist needs to get down to the business of convincing agnostics that they should not accept (1). One way to do so would be to convince agnostics that the compatibilist analysis of 'can' is correct. T&F acknowledge the problems with the simple conditional analysis of 'can', but perhaps they have a revived analysis in mind? Or a counterexample to (1) motivated independently of compatibilism?<sup>14</sup> This may not be necessary for making the case that (1) *begs the question* against the compatibilist, but it does seem to be necessary for doing the real work of convincing agnostics that that they should not accept (1).

### 3. Conclusion

I do not claim that the argument from FFP and this paper counts as a success according to van Inwagen's test: would it convince an audience of ideal agnostics? Who knows? But I would not therefore conclude that the argument is a "philosophical failure." Nor would I consider van Inwagen's Consequence Argument a failure for that reason. By deriving

---

<sup>14</sup> Cf. van Inwagen (1983) on his modal argument: "If the compatibilist wishes to refute these arguments—and, of course, nothing obliges him to do this—here is what he will have to do: he will have to produce some set of propositions intuitively more plausible than the premises of these arguments and show that these propositions entail compatibilism, or else he will have to devise arguments for the falsity of some of the premises employed in the present chapter, arguments that can be evaluated and seen to be sound independently of the question whether free will and determinism are compatible" (104).

incompatibilism from more general modal principles, these arguments clarify the logic and the costs of possible positions about the relation between freedom and modality.<sup>15</sup>

## References

- Chalmers, David J. (2012). *Constructing the World*. Oxford University Press.
- Fischer, John Martin and Garrett Pendergraft (2013). “Does the consequence argument beg the question?” *Philosophical Studies*, Vol. 166, No. 3, 575-595.
- Holliday, Wesley H. (2012). “Freedom and the Fixity of the Past,” *Philosophical Review*, Vol. 121, No. 2, 179-207.
- Horgan, Terence (1979). “‘Could’, Possible Worlds, and Moral Responsibility,” *The Southern Journal of Philosophy*, Vol. 17, Issue 3, 345-358.
- van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford University Press.
- van Inwagen, Peter (2006). *The Problem of Evil*. Oxford University Press.
- Lewis, David (1973). *Counterfactuals*. Basil Blackwell, Oxford.
- Nozick, Robert (1981). *Philosophical Explanations*. Harvard University Press.
- Spencer, Jack (2013). “Able to do the Impossible,” manuscript.
- Tognazzini, Neal A. and John Martin Fischer (2013). “Incompatibilism and the Past,” this volume. URL: <http://philpapers.org/archive/TOGIAT.pdf>
- Unger, Peter (1984). *Philosophical Relativity*. Oxford University Press.
- Vihvelin, Kadri (1995). “What Time Travelers Cannot Do.” *Philosophical Studies*, Vol. 81, 315-330.
- Williamson, Timothy (2002). *Knowledge and Its Limits*. Oxford University Press.

---

<sup>15</sup> I wish to thank Mark Crimmins for helpful discussion on the topic of this paper and Timothy Clarke for helpful comments on an earlier draft.