

## REPLY TO CHRISTOPHER HILL

By PETER VAN INWAGEN

**I**N MY BOOK *An Essay on Free Will* [3] and elsewhere, I have presented several arguments for the incompatibility of free will and determinism. The body of Christopher Hill's paper [1] is an examination and criticism of one of these arguments.<sup>1</sup> Hill's criticism is directed at the interaction between two facets of this argument: its sixth premiss

(6) J could not have rendered L false.

and the definition I provided in the course of the argument for the phrase 's can render p false'. In what follows, I will presume familiarity with my argument and with Hill's paper, and I shall assume that, in all of my references to premiss (6), this premiss is to be understood in the sense provided by my definition of 'can render false'.

The universal generalization of (6) — to all agents — is equivalent to incompatibilism.<sup>2</sup> I will, in the interests of brevity, summarize this fact by saying that (6) entails incompatibilism and incompatibilism entails (6). (This summary is strictly correct only on the assumption that if J could not have rendered L false, then no one could ever render L false. But perhaps this assumption is not likely to be very controversial.) In *An Essay on Free Will*, I took note of the fact that (6) entailed incompatibilism ([3], pp. 75–5). (More exactly, of the fact that if J could have raised his hand at T — a moment at which his hand was in fact unraised — and if determinism was true, then it followed that (6) was false.) And I concluded from this fact that any compatibilist who was not moved by my argument to recant his error would therefore want to reject (6) — and only (6), the other five premisses of the argument being obvious logical consequences of my description of J's situation. It is because the other premisses are in this sense trivial — together with the fact that the argument is a formally valid argument for the conclusion that if determinism is true, then J could not have raised his hand at T — that if J could have raised his hand at T, and if determinism is true, then it follows that (6) is false. Some have suggested that my use of a premiss that entails my conclusion convicts me of the fallacy of begging the question. As R. M.

<sup>1</sup> In *An Essay on Free Will*, the argument Hill discusses appears on pp. 56–78.

<sup>2</sup> Premiss (6) is the thesis that for any act that J could have performed but did not perform, it is false that the proposition that J performs that act and the prior state of the world together entail the denial of L. Incompatibilism may be formulated as the thesis that for any agent x, and for any act that x could have performed but did not perform, the non-performance of that act was undetermined; that is, it is false that L and the prior state of the world, together entail the proposition that x does not perform that act.

Chisholm is supposed once to have responded to a similar accusation, 'I seem to stand accused of the Fallacy of Affirming the Antecedent'.<sup>3</sup>

I am used to being charged with begging the question because I have employed a premiss that entails incompatibilism. If I understand Hill, his argument appeals to the reverse entailment: he accuses me of begging the question because I have employed a premiss that is entailed by incompatibilism. As I read his argument, the core of it is as follows. He points out that (6) is not a self-evident truth, and he goes on to contend that the only way to defend (6) is to appeal to the proposition he labels (9), which is equivalent to the universal instantiation of incompatibilism to J. At any rate, this is how *I* would describe Hill's argument. But he might not accept this description. *He* contends that his argument shows that I face a 'problem', which he characterizes as follows:

The problem, then, is this: van Inwagen's definition of 's can render *p* false' fails to sustain premiss (6) by itself. Further, if van Inwagen were to try to paper over the logical gap that separates (6) from the definition, he would have to assume an additional premiss [i.e., (9)] — a premiss that no compatibilist would grant. ([1], p. 53)

I find this passage puzzling. My definition of 'can render false' was not supposed to 'sustain' premiss (6) or any other proposition. It was, rather, supposed to be an explanation of what the term of art 'can render false' means. It is, of course, true that this phrase occurs (in the past tense, and negated) in premiss (6), and this fact establishes the only connection I intended to hold between the definition and the premiss: to understand the premiss you have to know the definition. Here is a similar and no doubt related point. I understand what it means to say that there is a 'logical gap' separating *propositions*: it means that a certain proposition can't be validly deduced from certain other propositions. But what does it mean to say that there is a logical gap separating a *definition* and a proposition?

At any rate, it is Hill's argument as I have formulated it above that I shall reply to. There are three things I wish to say in reply.

First, if one presents a formally valid argument for a certain conclusion, those who have hitherto rejected the conclusion must, if they are not converted by the argument and if they choose to reply to it, reject one or more of its premisses (or perhaps the disjunction of two or more of its premisses). Since there are no *proofs*

<sup>3</sup>There is certainly nothing dialectically improper in presenting an argument that has only one non-trivial premiss — or even only one premiss *tout court*. A philosophical argument, I take it, is supposed to uncover entailments that were not initially apparent. If I show that *p* entails *q* (or even that *p* is equivalent to *q*), I may thereby present a cogent argument for *q*, for if the entailment (or equivalence) was not initially apparent, my argument may thereby get people to see *q* in a new light. Perhaps *q* will look more plausible when seen in that light.

of interesting philosophical conclusions, such a response will always be an option. Since many philosophers are deeply committed to various philosophical theses — and compatibilism is certainly one of them — this option will often be taken. As I have said, (6) is the premiss of my argument that the unregenerate compatibilist will reject.

Secondly, I have never regarded it as my task to present an argument for incompatibilism all of whose premisses would be accepted by every compatibilist (or by most compatibilists or by typical compatibilists). This would be a quite impossible task, for the reasons touched on in the preceding paragraph. The much more modest task I have set myself can be perspicuously described in terms of a debate between me and a composite figure called 'the compatibilist'. (I ought to have said in the book what I am about to say here. I did not say it because it was so obvious to me that this is what I was up to that I did not explicitly formulate it, even to myself.) A debate implies an audience. Imagine that 'the compatibilist' and I are debating before an audience of people who are, so to speak, agnostics about the compatibility of free will and determinism. My object in this debate is not to convert the compatibilist to incompatibilism (although I should be pleased if he did see the light), and his object is not to convert me to compatibilism. Rather, each of us is trying to convert the members of the audience from agnosticism to, as the case may be, incompatibilism or compatibilism. My object, therefore, does not require me to find premisses that will be acceptable to the compatibilist. I *expect* him to reject one or more of my premisses. (He had better, if he wants to remain a compatibilist. I even *know* which one he will reject — at least if he knows his business.) What my object requires is, rather, that I articulate my reasons for being an incompatibilist in such a way as to make them accessible to the audience of agnostics, that I attempt to undermine the appeal of the compatibilist's arguments for compatibilism, and that I do what I can to counter his attempts to undermine the appeal of my arguments. It is in the nature of such a debate that if I have any reasons for being an incompatibilist that, in my judgment, would not appeal to the agnostics, I shall not bother to mention them. I am interested only in presenting premisses that I judge it likely that the audience, or some in the audience, will accept. I think that (6) is such a premiss. I have never, in my published work on the problem of free will and determinism, tried to justify (6) by deducing it from some other proposition — such as Hill's (9) — for it has always seemed to me that (6) was pretty plausible just as it stood.<sup>4</sup> (I, of course, accept (9). But I do not think that this proposition is any closer to being evidently true than is (6), and I expect that the most of the members of the audience think so, too. There would, therefore, be no point in my trying to support (6) by appealing to (9).)

So much for the role of premiss (6) in the debate. What about my definition of 'can render false'? It occurs for the following reason. At a certain point in the debate, prior to the introduction of that definition, the compatibilist suggests that there is no single sense of 'can render false' in which all of the premisses of my argument that contain this phrase are true.<sup>5</sup> I introduce the definition and contend that they all look pretty good to me when 'can do otherwise' is understood in that sense.<sup>6</sup> The compatibilist replies that (6) is false when it is understood according to the terms of this definition. Well, each of us has now made his case. I am content to let each of the agnostics decide for himself which of us is right. But there is something more to be said about the plausibility of (6), which brings me to . . .

Thirdly, although I have not attempted to deduce (6) from some more fundamental premiss or set of premisses, I have adduced certain considerations that I believe render it quite plausible (cf. [3], pp. 75–8). I shall briefly rehearse them. Let us replace 'L' in (6) with some particular law of nature. ('L', remember, is the conjunction of all laws of nature.) Let us pick 'No two bodies have a relative velocity greater than that of light'. The resulting thesis is plausible indeed. According to the definition I have provided for 'can render false',<sup>7</sup> this thesis is equivalent to

There is no action A such that (e) it is within J's power to perform A and (f) there is no logically possible world *w* such that (i) J performs A in *w*, (ii) the state of *w* up to the time at which J performs A is exactly the same as the state of the actual world up to that time, and (iii) it is true in *w* that no two bodies have a relative velocity greater than that of light.

Consider the clause (f). What actions might satisfy the condition laid down by this clause? Here are three: drawing a round square, moving one's left hand faster than light, and (assuming that in the actual world J is an inhabitant of the earth) visiting Arcturus tomorrow. The third is the most interesting, since all three of the

<sup>4</sup>Hill says (p. 52) that (6) '... loses its appearance of correctness when it is understood in terms of van Inwagen's definition'. But to *whom* does it no longer appear correct? It still appears correct to *me*. I hope, and, I believe, not without some reason, that it will still appear correct to some of the members of the audience.

<sup>5</sup>See, for example, David Lewis, 'Are We Free To Break the Laws?' [2].

<sup>6</sup>I have cited Lewis's paper (in the previous note) because it is, by far, the best paper I am aware of in which the charge mentioned in the text is made. It is sometimes supposed that I framed my definition of 'can render false' with the intention of blocking the argument of 'Are We Free To Break the Laws?'. In fact, however, when Lewis's paper reached me I was making the final revisions of the manuscript of *An Essay on Free Will*, and the definition was by then already 'in place'. The definition is the product of conversations I had been having with Mark Heller, who was then a graduate student.

<sup>7</sup>I am here using a formulation of the definition taken from Hill's paper (p. 51).

sub-clauses of (f) have to be taken into account in a demonstration that this action satisfies (f). But, of course, J *cannot* visit Arcturus tomorrow. And a sufficient reason for J's inability to visit Arcturus tomorrow is provided by the following two facts: the laws of nature do not allow the distance between J and Arcturus to decrease by more than one light-day in a given twenty-four-hour period; J and Arcturus are today almost fourteen thousand light-days apart. Therefore, the action *visiting Arcturus tomorrow* does not satisfy clause (e) of the sentence displayed above. And isn't it evident that *no* action that satisfies (f) will satisfy (e)? (Certainly *drawing a round square* and *moving one's left hand faster than light* do not satisfy (e).) Hence, the result of substituting 'No two bodies have a relative velocity greater than that of light' for 'L' in (6) is obviously true.

Now suppose that we were to substitute some other particular law of nature for 'L' in (6), and were to attempt to devise an argument parallel to the above argument. Doesn't it seem plausible to suppose that, for the case of any law of nature, we could devise an argument parallel to the argument of the preceding paragraph and no less cogent? It is hard to see why we shouldn't be able to do this. Suppose we were to substitute the conjunction of *two* laws for 'L'. Doesn't it seem plausible to suppose that, for the case of the conjunction of any two laws of nature, we could devise an argument parallel to the argument of the preceding paragraph and no less cogent? Again, it is hard to see why we shouldn't be able to do this. And what difference would it make if we considered not a particular law of nature, or a conjunction of a few laws, but the conjunction of *all* laws?

The preceding paragraph does not contain an argument for (6). It is, after all, nothing more than a series of questions — like the well-known series that culminates in, 'Whence, then, is evil?'. This series of questions, however, does, to my mind, lend a great deal of plausibility to (6).

Has the compatibilist a way of replying to this series of questions? Of course. The compatibilist will say that I have chosen my primary example of a law very carefully, since this law is not one that could easily be imagined as figuring in a causal explanation of any human action.<sup>8</sup> But, the compatibilist will say, if you had just begun to whistle — which you did not — and if the world is deterministic, and if you were able to whistle, then (given your definition of 'can render false') it follows that there is a law, one that figures in the explanation of why you did what you were doing a moment ago (something incompatible with whistling),

<sup>8</sup> David Lewis made this point in correspondence, after he had sent me a manuscript of 'Are We Free To Break the Laws?' and I had responded with a letter containing the definition of 'can render false' that was to appear in *An Essay on Free Will*, and an argument similar to that presented above in the text.

such that you were able to render that law false. Or at least it follows that there is a set of laws that figure in the explanation of your action, such that you were able to render the conjunction of the members of this set false. My response — the attentive reader will have anticipated it — is that this does indeed follow, and, since it's wholly implausible to suppose that anyone has the power to render a law of nature false (either in the sense I have given to 'can render false', which is a reasonable one, or in any other reasonable sense), the conclusion that one ought to draw is that either the world is not deterministic or else I was not able to whistle. (I should remark that although it is true that my primary example of a law is not one that could easily be imagined as figuring in a causal explanation of any human action, and although I may have got some illegitimate dialectical advantage by using a law that had that feature as an example, I at any rate did not choose it *because* it had that feature; I chose it because it was a law that constrained human action in a clear and uncontroversial way.)

This response will not convert all compatibilists to incompatibilism. Perhaps it will not even convert very many of them. (It *has* convinced some.) But I am not primarily interested in converting compatibilists. I have my eye on the audience.

*Syracuse University,  
Syracuse, NY 13244-1170, USA*

#### REFERENCES

- [1] Christopher Hill, 'Van Inwagen on the Consequence Argument', *Analysis* 52 (1992) 49–55.
- [2] David Lewis, 'Are We Free to Break the Laws?', *Theoria* 47 (1981) 112–21; reprinted in Lewis's *Philosophical Papers* Volume II (New York: Oxford University Press, 1986).
- [3] Peter van Inwagen, *An Essay on Free Will* (Oxford: Oxford University Press, 1983).

### URBACH ON THE LAWS OF NATURE

*By* CHRISTOPHER HITCHCOCK

PETER URBACH [4] has put forward a subjectivist account of laws of nature. He takes as his point of departure the following two observations: (1) one of the primary tasks of an account of the laws of nature is to explain the intuitive difference between law-