TERENCE HORGAN

COMPATIBILISM AND THE CONSEQUENCE ARGUMENT

(Received 20 February, 1984)

1

Peter van Inwagen, in an influential paper and again in a recent book, has propounded an important argument for the incompatibility of free will and determinism.[1] Of the various replies that have appeared in response to his original paper, perhaps the most incisive is by David Lewis.[2] Although van Inwagen does not discuss Lewis in his new book (which was already at press by the time the paper appeared), he does elaborate upon his own original argument in a way which suggests a response to Lewis's critique. In this paper I shall set forth, and then evaluate, that response.

van Inwagen points out that his argument represents one way of refining the following line of reasoning, which he calls the *Consequence Argument*:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (ETW, p. 16)

He presents two other elaborations of the Consequence Argument in his book, and observes that all three versions probably stand or fall together. Other defenders of one or another version of the Consequence Argument include Carl Ginet, James Lamb, and David Wiggins.[3] Michael Slote has described well the deep family resemblances among the various formulations, and he too has suggested that the different versions probably all stand or fall together.[4] Thus, if Lewis is correct in his critique of the specific version we shall consider here, then the other versions are probably in trouble too. Conversely, if van Inwagen's most recent discussion provides the basis for an adequate refutation of Lewis, then Lewis-style objections to the other versions can probably be refuted as well.

Philosophical Studies **47** (1985) 339–356. 0031–8116/85.10

2

I begin with a summary of van Inwagen's argument and Lewis's reply. van Inwagen asks us to imagine the following scenario. There once was a judge, J, who had only to raise his hand at a certain time, T, to prevent the execution of a certain criminal. J refrained from doing so at T, so that the criminal was subsequently executed. J was unbound, uninjured, and free from paralysis. His decision came about only after a period of calm, rational, and relevant deliberation. He was psychologically normal. And he was not under the influence of drugs, alcohol, or anything of that sort.

Van Inwagen uses 'T_0' to denote some instant of time prior to J's birth, 'P_0' to denote a proposition which expresses the total intrinsic state of the world at T_0, 'P' to denote a proposition which expresses the total intrinsic state of the world at T, and 'L' to denote the conjunction into a single proposition of all the laws of physics. His overall argument consists of the following truth-functionally valid formal argument, together with an accompanying commentary in defense of the six premises.[5]

(1) If determinism is true, then the conjunction of P_0 and L entails P.

(2) It is not possible that J have raised his hand at T and P be true.

(3) If (2) is true, then if J could have raised his hand at T, J could have rendered P false.

(4) If J could have rendered P false, and if the conjunction of P_0 and L entails P, then J could have rendered the conjunction of P_0 and L false.

(5) If J could have rendered the conjunction of P_0 and L false, then J could have rendered L false.

(6) J could not have rendered L false.

$\therefore$ (7) If determinism is true, J could not have raised his hand at T.

Exactly analogous reasoning can be used to argue that if determinism is true, then nobody *ever* can act otherwise than he does act.

In his original paper van Inwagen does not attempt to explicate the idiom 'S can render [could have rendered]... false', but instead treats it as a fairly natural extension of our ordinary used of 'can' and 'could' as appended to action-verbs. Lewis, however, does undertake to give an explicit meaning to this term of art. In fact he considers two possible meanings, and he argues that neither will serve van Inwagen's purposes.

Lewis offers the following preliminary definition: an event *would falsify* a proposition iff, necessarily, if that event occurs then that proposition is false. Then come the two alternative definitions of 'could have rendered false'. First, an agent *could have rendered a proposition false in the weak sense* iff the agent was able to do something such that, if he did it, the proposition would have been falsified (though not necessarily by his act, or by any event caused by his act). Second, an agent *could have rendered a proposition false in the strong sense* iff he was able to do something such that, if he did it, the proposition would have been falsified either by his act itself or by some event caused by his act.

If we take the weak sense throughout the above derivation, says Lewis, then the compatibilist can plausibly deny Premise 6. He can claim that J was able to do something (e.g., raise his hand at T) such that if he did it, some event or other would have falsified a law. The relevant event would be some relatively minor law-violation just prior to J's (counterfactual) act, just enough of a violation to smoothly graft this act onto the actual world's past. (This is what Lewis calls a "divergence miracle".) Lewis stresses that when one claims that J could have rendered L false in the weak sense, one does not thereby commit oneself to the fantastic claim that J could have *broken* a natural law. For, to break a law would be to do something such that, if one did it, then either one's act itself or else some event caused by the act would falsify L. I.e., to break a law would be to render L false in the *strong* sense.

If we take the strong sense of 'could have rendered false' throughout van Inwagen's derivation rather than the weak sense, says Lewis, then the compatibilist can plausibly deny Premise 5. Although J could indeed have rendered the conjunction of P_0 and L false in the strong sense (say, by raising his hand at T), and although J could not have rendered P_0 false in the strong sense (or even in the weak sense), nevertheless J could not have rendered L false in the strong sense. For, if J had raised his hand at T then L would have been falsified not by J's act or by any event caused by that act, but instead would have been falsified by a prior divergence miracle.

This last point is clarified by examining the assumptions that lie behind van Inwagen's Premise 5. It is reasonably clear from his discussion that he infers Premise 5 from the following two propositions, each of which he evidently takes to be uncontroversial (and, indeed, analytic).[6]

(A) If S can render false the conjunction of two propositions p and q, then either S can render p false or S can render q false.

(B) If p is a true proposition that concerns only states of affairs that obtained before S's birth, then S cannot render p false.

But principle (A), far from being uncontroversial, is entirely unwarranted under the strong sense of 'can render false'. For, the inference from (C) to (D) is invalid:

(C) Necessarily, if the event E occurs then the conjunction of the propositions p and q is false.

(D) *Either* necessarily, if the event E occurs then p is false; *or* necessarily, if the event E occurs then q is false.

Thus an event E might well falsify a conjunctive proposition (p and q) without falsifying either p or q, and hence an agent sometimes might well be able to render a conjunctive proposition false in the strong sense even if he cannot render either conjunct false in the strong sense.[7] van Inwagen's judge is a case in point, relative to the propositions P_0 and L. The judge can raise his hand at T, and under determinism this act would falsify (P_0 and L); but neither the act itself nor any of its effects would falsify either L or P_0. (L would indeed be falsified, but by a prior divergence miracle.)

So under Lewis's analysis, van Inwagen's argument fails under either reading of 'could have rendered false'. Its *prima facie* plausibility rests largely upon our tendency to equivocate between the two readings.

3

There is controversy among philosophers about the way counterfactuals in general, and counterfactuals about human agency in particular, behave under determinism. Some philosophers, following Lewis, maintain that if the antecedent of a given counterfactual were true, then (a) the past would have been the same as it was in the actual world, up until just prior to the time of the antecedent, and (b) a minor last-moment divergence miracle would have occurred, just enough of a miracle to guarantee the truth of the antecendent.[8] Others, however, maintain that if the antecedent were true, then the actual world's laws would have been unviolated but the past would have differed slightly, at each past moment of time, from the actual world's past.[9] Let us call the former approach the *miraculous analysis* of counterfactuals, and the latter approach the *nonmiraculous analysis.* And let us say that compatibilists

who advocate the miraculous analysis are *divergence-miracle compatibilists*, whereas those who advocate the nonmiraculous analysis are *altered-past compatibilists*.

An altered-past compatibilist who wishes to reply to van Inwagen's argument also can avail himself of Lewis's distinction between the two senses of 'could have rendered false'. But rather than attacking Premise 6, he will claim that Premise 5 is false under both the weak sense and the strong sense of this phrase. If we take the weak sense, he will say, then claim (B) above is false; for, J can do something (e.g., raising his hand at T) which is such that, if he did it, then P_0 would have been false. And if we take the strong sense, then claim (A) is false; for, although J can render (P_0 and L) false in the strong sense, J cannot render P_0 false in the strong sense and he cannot render L false in the strong sense. So according to the altered-past compatibilist, either (A) or (B) is false; and Premise 5 is false either way.[10]

I shall remain neutral here about the vexing problem of how counterfactuals behave under determinism. Rather than taking a stand on this controversial matter, I shall conduct the subsequent discussion disjunctively. I.e., I shall examine the crucial issues from the perspective of each kind of compatibilist in turn.

4

In his recent book van Inwagen repeats and elaborates his original argument, but unfortunately he does not discusses Lewis's rejoinder. However, he does address the question of the meaning of 'could have rendered false', and my present concern is to examine his discussion of this matter in light of Lewis's critique.

He first explicitly takes up the meaning of 'could have rendered false' in his 'Reply to Narveson', a paper that appeared after the original one and prior to the book.[11] There he offers this preliminary definition: a state of affairs *entails the falsity* of a proposition p iff it is not possible that this state of affairs obtain and p be true. Then he defines the key phrase this way:

> An agent S *can render false* the proposition p iff: either p is false or, if p is true, then there is some state of affairs A such that (a) S can (i.e., has it within his power to) bring about A, and (b) A entails the falsity of p.[12]

This definition is essentially equivalent to Lewis's definition of 'can render

false in the strong sense'. Accordingly, it is of no help in evading Lewis's criticism. For, now the compatibilist can plausibly deny Premise 5 by denying claim (A), as explained above.

(Interestingly, in the paper where he proposes this definition, van Inwagen defends Premise 5 by arguing at some length in favor of (B). He simply takes (A) for granted. Yet (A) is the really dubious claim, under his definition.)

In his new book, however, he argues that the above characterization of 'can render false' is not the appropriate one. He says:

> Let us suppose that in 1550 Nostradamus predicted that the Sphinx would endure till the end of the world. And let us suppose that this prediction was correct and, in fact, that *all* Nostradamus's predictions were correct. Let us also suppose that it was within Gamal Abdel Nasser's power to have the Sphinx destroyed. Then, I should think, it was within Nasser's power to render false the proposition that all Nostradamus's predictions were correct. But this would not be the case according to the [above definition], since it is possible in the broadly logical sense that Nasser have had the Sphinx destroyed and yet all Nostradamus's predictions have been correct. (ETW, p. 67)

On this basis of this example, he proposes to define '*S* can render *p* false' as follows:

> It is within *S*'s power to arrange or modify the concrete objects that constitute his environment in some way such that it is not possible in the broadly logical sense that he arrange or modify those objects in that way and the past have been exactly as it in fact was and *p* be true. (ETW, p. 68)

This is a considerably more liberal definition – that is, a considerably *weaker* definition. Also, the definition seems preferable to either of Lewis's, on one score at least: *viz*., it allows that Nasser has the power to falsify the proposition that all of Nostradamus's predictions are true. (Let us call this proposition '*N*'.) Lewis, on the other hand, evidently must deny that Nasser can falsify *N* in either the strong sense or the weak sense. For, both of Lewis's definitions employ the same stringent notion of an *event's* falsifying a proposition: an event *E* would falsify a proposition *p* iff *E*'s occurrence *strictly implies* that *p* is false.

Given that van Inwagen's definition seems preferable to either of Lewis's in its handling of the Nostradamus example, it now appears that Lewis's critique is flawed by his implausibly strong characterization of an event's falsifying a proposition. Can van Inwagen exploit this problem as a way of evading Lewis's criticisms, or will the problems simply re-appear in somewhat altered form? This is the key question.

5

Let us consider whether Lewis's definitions are really incapable of accomodating the case of Nostradamus. One might try arguing as follows for the claim that Nasser actually can render false, in the strong sense, the proposition that all of Nostradamus's predictions are true:

There is no single proposition expressed by the sentence 'All of Nostradamus's predictions are true'. Rather, we should distinguish two propositions, N_1 and N_2, either of which can be expressed by this sentence. Let N_1 be the proposition van Inwagen has in mind: a proposition that is true at a possible world w iff all of Nostradamus's predictions in w are true at w. But there is also N_2, a proposition that is true at a world w iff all of Nostradamus's *actual-world* predictions are true at w. With his distinction at hand, it turns out that there is one perfectly legitimate sense in which Nasser could have rendered false, in the strong sense, the proposition that all of Nostradamus's predictions are true. To wit: he could have rendered N_2 false in the strong sense. Thus there is really nothing wrong with Lewis's two definitions after all.

Originally I was attracted by this line, but I now think it is a red herring. For, I don't think that Nasser really *can* render N_2 false in the strong sense, under Lewis's definition.

In order to be able to render N_2 false in the strong sense, Nasser must be able to perform some act A such that either A itself or one of its effects would falsify N_2. Now, obviously the event which supposedly has this feature is the destruction of the Sphinx. But is the destruction of Sphinx, occurring at a time shortly after Nasser's order that the Sphinx be destroyed, an event which *strictly implies* the falsity of N_2 – as required by the Lewis's definition of an event's falsifying a proposition? I think not.

To see why not, suppose that when Nostradamus predicted that the Sphinx would last until the end of the world, he meant that it would last until the end of human civilization – not until the end of time. Consider a possible world with the following features: (a) Nostradamus makes all of his actual-world predictions; (b) Nasser orders the destruction of the Sphinx; (c) shortly thereafter, the Sphinx is destroyed by Nasser's men, as a result of Nasser's order; and (d) a split second after the Sphinx is destroyed, the Martians destroy the entire earth and everybody living on it. In this world the Sphinx-destruction occurs shortly after Nasser's order and as a result of his order, and yet Nostradamus's prediction about the Sphinx is not false. Since such a world exists, the proposition that the event in question occurs does not strictly imply that Nostradamus's prediction is false. Hence this event would not falsify Nostradamus's prediction, which means (a) that

Nasser cannot render this prediction false in the strong sense, and hence (b) that Nasser cannot render either N_1 or N_2 false in the strong sense.

So it appears that Lewis's definitions cannot accomodate the Nostradamus case after all. Furthermore, it seems entirely natural to say that Nasser can render Nostradamus's prediction false, and hence that he can render false the proposition that all of Nostradamus's predictions are true. So van Inwagen seems justified in proposing a definition of 'can render false' which will accomodate these intuitions.

6

The problem with Lewis's two definitions of 'can render false' is that they rely on too stringent a notion of an event's falsifying a proposition. Similarly with van Inwagen's earlier definition of 'can render false', which employs the (essentially equivalent) notion of a state of affairs *entailing the falsity* of a proposition.[13] Van Inwagen's revised definition handles this problem in a seemingly natural way: in effect, it builds into the definition of 'can render false' the idea that the falsifying of a proposition by a state of affairs is a context-dependent matter. The context, of course, is provided by the actual past: an event or state of affairs falsifies a proposition iff it's not simultaneously possible for (a) that event to occur, (b) that proposition to be true, and (c) *the past to remain as it actually was*.

This general approach really should be altered somewhat, in order to allow for any differences from actuality that would accompany the given event or state of affairs – *viz*., either a last-moment divergence miracle or a moderately-altered entire past. The relevant context, relative to a given event or state of affairs, should include not the *entire* actual past, but rather the past that *would have* existed had the event occurred.[14]

Furthermore, we really need to consider more than just the past if we want to accomodate our intuitive idea that Nasser can render false Nostradamus's prediction that the Sphinx will last until the end of civilization. For, presumably there exists a possible world with the following features: (a) Nostradamus makes all of his actual-world predictions; (b) Nasser orders the destruction of the Sphinx; (c) shortly thereafter, the Sphinx is destroyed by Nasser's men, as a result of his order; (d) a split second later, the Martians destroy the entire earth and everybody living on it; and (e) the past, prior to Nasser's order, is the same as our actual world's past, except for whatever

differences there would have been had Nasser ordered the destruction of the Sphinx. (In this world the Martians don't exist in the past; after Nasser's order, they literally appear from nowhere. Let's not forget just now broad 'broadly logical possibility' really is.) In order to prevent such a world from undermining the claim that Nasser can render Nostradamus's prediction false, I think we should let the relevant context of Nasser's action include not merely the entire *past* that would have obtained had Nasser ordered the destruction of the Sphinx, but rather the entire *past, present, and future* that would have obtained.

So let us say that an event *E would falsify in the broad sense* a proposition *p* iff there is a true proposition *q* such that (i) if *E* were to occur then *q* would still be true, and (ii) necessarily, if *E* occurs and *q* is true then *p* is false. And now, using this liberalized definition in place of Lewis's more stringent definition of an event's falsifying a proposition, we can adopt almost verbatim Lewis's definitions of 'can render false in the weak sense' and 'can render false in the strong sense'. Let us say that an agent *can render a proposition false in the weak and broad sense* iff the agent is able to do something such that, if he did it, the proposition would be falsified in the broad sense (though not necessarily by his act, or by any event caused by his act). And let us say that an agent *can render a proposition false in the strong and broad sense* iff he is able to so something such that, if he did it, the proposition would be falsified in the broad sense either by his act itself or by some event caused by his act.

The important question, for our purposes, is whether these revised definitions will help van Inwagen's argument for the incompatibility of free will and determinism. How does the argument fare if we use 'can render false in the weak and broad sense', or 'can render false in the strong and broad sense'? In particular, what is the status of the crucial premises 5 and 6?

Here we must bifurcate the discussion, in order to consider the matter both from the perspective of the divergence-miracle compatibilist and also from the perspective of the altered-past compatibilist.

7

Let us first adopt the viewpoint of the altered-past compatibilist. He will claim that if we use 'can render false in the weak and broad sense' throughout the argument, then Premise 5 is false because principle (B) above is false.

For, the judge J can do something (*viz*., raising his hand at T) such that, if he did it, then the proposition P_0 would have been falsified in the broad sense by some event or other; in particular, P_0 would have been falsified by an event occurring at T_0 which did not occur in the actual world, and which was a remote causal ancestor of the judge's raising his hand at T.

What if we take 'can render false in the strong and broad sense' throughout the argument? Now the altered-past compatibilist cannot say, as he did earlier in relation to Lewis's definition of 'can render false in the strong sense', that principle (A) is false with respect to the judge J. For, if determinism is true then J's act of raising his hand at T not only would falsify (P_0 and L) in the broad sense, but it also would falsify P_0 in the broad sense. (The proof is as follows. Assuming that determinism is true, L is a true proposition q such that (i) if J's raising his hand at T were to occur then q would still be true, and (ii) necessarily, if J's raising his hand at T occurs and q is true, then P_0 is false. Hence J's raising his hand at T is an event which would falsify P_0 in the broad sense.)

On the other hand, the altered-past compatibilist can plausibly reject (B), relative to 'can render false in the strong and broad sense'. We have just established that if determinism is true, and if the nonmiraculous analysis of counterfactuals is correct, then J's hand-raising *itself* – and not merely some event that occurs at the remote past time T_0 – is an event that would falsify P_0 in the strong and broad sense. If we keep this fact well in mind, and if we assume that the nonmiraculous analysis of counterfactuals is correct, then it is not at all implausible to say that J can render P_0 false in the strong and broad sense. For, J can raise his hand at T, and this act itself is now being counted as a P_0-falsifying event. It would be outrageous, of course, to claim that J can *causally influence* events in the remote past. But we are saying nothing so offensive when we assert that J can render P_0 false in the strong and broad sense. On the contrary, essentially all we are saying is that J can do something that he is causally determined not to do; and it is no surprise to learn that the compatibilist is committed to *that*.

So the altered-past compatibilist will claim, with plausibility, that principle (B) is false in relation to each of our two recent renderings of 'can render false'; thus Premise 5 is false either way. Hence neither of these liberalized definitions of 'can render false' will serve van Inwagen's purposes, if counterfactuals receive the nonmiraculous analysis.

8

Let us now adopt the viewpoint of the divergence-miracle compatibilist. He will claim that if we use 'can render false in the weak and broad sense' throughout the argument, then Premise 6 is false. J can do something (*viz.*, raise his hand at T) such that, if he did it, then the proposition L would have been falsified by some event or other. In particular, it would have been falsified by a prior divergence miracle.

What if we take 'can render false in the strong and broad sense' throughout the argument? Now the divergence-miracle compatibilist cannot say, as he did earlier in relation to Lewis's definition of 'can render false in the strong sense', that Premise 5 is false by virtue of the falsity of Principle (A). On the contrary, Premise 5 is now true. For, suppose that J could have rendered false, in the strong and broad sense, the conjunction of P_0 and L – say by performing some act A (e.g., the act of raising his hand at T) that is incompossible with the proposition P which describes the total instrinsic state of the world at T. Now if we assume determinism, it is necessarily true that if J performs A and P_0 is true, then L is false. (This is because (P_0 and L) entails P, and A is incompossible with P.) Furthermore, if J had performed A then P_0 still would have been true. (A divergence miracle would have preceded A, but the remote past as described by P_0 would be the same as in the actual world.) Hence J's performing A at T would falsify L, in the strong and broad sense. Therefore, if J can render (P_0 and L) false in the strong and broad sense, then J can render L false in the strong and broad sense.

(In this defense of Premise 5, we needed to assume determinism. So Premise 5 should be rewritten this way:

> If determinism is true, then if J could have rendered the conjunction of P_0 and L false, then J could have rendered L false.

But this change does not affect the truth-functional validity of van Inwagen's derivation.)

But although the divergence-miracle compatibilist cannot deny Premise 5, relative to 'can render false in the strong and broad sense', he can plausibly reject Premise 6. We have just established that if determinism is true, and if the miraculous analysis of counterfactuals is correct, then J's hand-raising *itself* – and not merely some event that occurs at the remote past time T_0 – is an event that would falsify L in the strong and broad sense. If we keep this

fact well in mind, and if we suppose that the miraculous analysis of counterfactuals is correct, then it is not at all implausible to say that J can render L false in the strong and broad sense. For, J can raise his hand at T, and this act itself is now being counted as an L-falsifying event. It would be outrageous, of course, to claim that J can do something such that either the act itself, or one of its effects, *violates* a law – that is, falsifies a law in Lewis's original strict sense of event-falsification. But we are saying nothing so offensive when we assert that J can render L false in the strong and broad sense. On the contrary, essentially all we are saying is that J can do something that he is causally determined not to do; and it is no surprise to learn that the compatibilist is committed to *that*.

So the divergence-miracle compatibilist will claim, with plausibility, that Premise (6) is false in relation to each of our two recent readings of 'can render false'. Hence neither of these liberalized definitions of 'can render false' will serve van Inwagen's purposes, if counterfactuals receive the miraculous analysis.

9

Lewis's concept of an event's falsifying a proposition may well be overly stringent; the Nostradamus example seems to demonstrate this. But we now see that our proposed successor-concept, the notion of an event's falsifying a proposition in the broad sense, is too inclusive to accomodate the intuitively natural claim that J's raising his hand at T is neither (a) an event which would falsify P_0 under the nonmiraculous analysis of counterfactuals, nor (b) an event which would falsify L under the miraculous analysis of counterfactuals. And as we have seen, the effect of this inclusiveness is that 'can render false in the strong and broad sense' will not serve van Inwagen's purposes. (Nor will 'can render false in the weak and broad sense'; for, this locution fares just the same as Lewis's original 'can render false in the weak sense'.)

I myself have little idea how we might frame a single definition of event-falsification under which both (a) the destruction of the Sphinx falsifies Nostradamus's prediction, and yet (b) J's hand-raising at T does not falsify either P_0 or L. The concept of event falsification turns out to be very elusive indeed. But even if such a definition can be found it won't help van Inwagen's argument, because some important general morals can be extracted from the above discussion.

If the nonmiraculous analysis of counterfactuals is correct, then the morals are these.[15] First, any definition of event-falsification that is broad enough to make principle (A) true, under the corresponding definition of 'can render false in the strong sense', will be a definition which classifies J's raising his hand at T as a P_0-falsifying event; so principle (B) will not be plausible, under that definition of 'can render false in the strong sense; and hence Premise 5 will not be plausible either. Second, any definition of an event's falsifying a proposition that is stringent enough to make principle (B) plausible, under the corresponding definition of 'can render false in the strong sense', will be a definition which precludes J's raising his hand at T from counting as a P_0-falsifying event (or an L-falsifying event), even though this act certainly will still count as a (P_0 and L)-falsifying event; so principle (A), and likewise Premise 5, will be false under that definition of 'can render false in the strong sense'. And third, regardless of how one defines event-falsification, principle (B), and likewise Premise 5, will be false under the corresponding *weak* sense of 'can render false'.

If, on the other hand, the miraculous analysis of counterfactuals is correct, then the morals that emerge from the above discussion are these. First, any definition of event-falsification that is broad enough to make Premise 5 true, under the corresponding definition of 'can render false in the strong sense', will be a definition which classifies J's raising his hand at T as an L-falsifying event; so Premise 6 will not be plausible, under that definition of 'can render false in the strong sense'. Second, any definition of an event's falsifying a proposition that is stringent enough to make Premise 6 plausible, under the corresponding definition of 'can render false in the strong sense', will be a definition which precludes J's raising his hand at T from counting as an L-falsifying event (or a P_0-falsifying event), even though this act certainly will still count as a (P_0 and L)-falsifying event; so principle (A), and likewise Premise 5, will be false under that definition of 'can render false in the strong sense'. And third, regardless of how one defines event-falsification, Premise 6 will be false under that corresponding *weak* sense of 'can render false'.

So the upshot is that Lewis's essential criticism cannot be evaded by definitional maneuvering: there is no single definition of 'can render false' that will serve van Inwagen's purposes. This conclusion should be reassuring to compatibilists, because arguments like van Inwagen's are perhaps the strongest yet provided by the incompatibilist camp.[16]

APPENDIX

The argument we have been examining is the first of three versions of the Consequence Argument presented in van Inwagen's book. I shall briefly explain how the above discussion can be transferred to the other two versions.

The second version may be paraphrased as follows. If free will exists, then at least one person 'has access' to at least one possible world other than the actual world. But nobody has access to any possible world in which L is not true, and nobody has access to any possible world whose total intrinsic state at each moment of time is different from the actual world's total intrinsic state at that time. So if determinism is true then nobody has access to any possible world other than the actual world, and hence free will does not exist.

It is now clear how a compatibilist can respond to this argument. An altered-past compatibilist can plausibly deny the premise that nobody has access to any possible world which differs somewhat, at each moment of time, from the actual world. He can claim, on the contrary, that if determinism is true, then sometimes an agent can do something which is such that if he did it then his act would have been preceded by a sequence of minor differences from actuality, backward throughout time. This claim is not to be confused with the incredible claim that sometimes an agent can causally influence the past.

A divergence-miracle compatibilist, on the other hand, can plausibly deny the premise that nobody has access to any possible world in which L is not true. He can claim, on the contrary, that if determinism is true, then sometimes an agent can do something which is such that if he did it then his act would have been preceded by a divergence miracle. This claim is not to be confused with the outrageous claim that agents can sometimes perform acts which are miracles themselves or which cause miracles.

Van Inwagen's third argument employs a modal operator, 'N'. 'Np' is to be rendered in English this way: 'p, and no one has, or ever had, any choice about whether p'. He adopts the following two plausible-seeming inference rules concerning this operator:

(α) $\quad \Box p \vdash Np$

(β) $\quad N(p \supset q), Np \vdash Nq$

With these rules at hand, he reasons as follows. Let 'P_0', 'L', and 'P' now be used as abbreviations for sentences, rather than as names of propositions.

'P_0' goes proxy for a sentence expressing a proposition about the total intrinsic state of the world at some instant in the remote past, and 'P' can be replaced by any true sentence one likes. Now if determinism is true, then it follows that

(1) $\Box(P_0 \text{ and } L. \supset P)$

is true. From (1) we may deduce

(2) $\Box(P_0 \supset (L \supset P))$

by elementary modal and sentential logic. Applying rule (α) to (2), we have:

(3) $N(P_0 \supset (L \supset P))$.

We now introduce a premise:

(4) NP_0.

From (3) and (4) we have by rule (β):

(5) $N(L \supset P)$.

We introduce a second premise:

(6) NL

Then, from (5) and (6) by (β):

(7) NP.

Thus, if determinism is true then no one ever has any choice about anything.

Now, clearly there are various ways one might construe the locution 'has a choice about whether', just as there are various ways one might contrue the notion 'can render false'. Hence my remarks at the end of section 9 are applicable, *mutatis mutandis*, to this argument. Do we interpret lines 4 and 6 in such a way that they entail that no one can, or ever could, bring about a past-falsifying event or law-falsifying event – where J's raising his hand at T counts (under determinism) as a past-falsifying event under the nonmiraculous analysis of counterfactuals, and as a law-falsifying event under the miraculous analysis of counterfactuals? If so, then the altered-past compatibilist can plausibly deny line 4, just as he can plausibly deny Premise 5 of the earlier argument by denying principle (B). And the divergence-miracle compatibilist

can plausibly deny line 6, just as he can plausibly deny Premise 6 of the earlier argument.

Or rather, do we interpret lines 4 and 6 in such a way they only entail that no one can, or ever could, bring about an event that would falsify a proposition about the past, or a law, in some suitably *stringent* sense of event-falsification? If so, then the compatibilist can plausibly claim that the operator 'N' is non-agglomerative, just as 'cannot render false in the strong sense' is non-agglomerative under any definition of event-falsification which precludes J's hand-raising at T from being a past-falsifying or a law-falsifying event. Specifically, one can argue that since J's raising his hand at T would falsify the conjunctive proposition (P_0 and L) without falsifying either P_0 or L, under a suitably stringent sense of event-falsification, it is therefore correct to say that under the corresponding interpretation of the operator 'N', the sentence '$N(P_0$ and $L)$' is false even though the sentences 'NP_0' and 'NL' are both true. Hence 'N' is non-agglomerative.[17]

But the inference rule (β) rests upon the principles of agglomerativity and closure-under-entailment: agglomerativity takes us from 'Np' and '$N(p \supset q)$' to '$N[p$ and $(p \supset q)]$': and closure then yields 'Nq'.[18] Thus, if 'N' is non-agglomerative then (β) is invalid.

So the compatibilist can claim, with justification, that there is no single construal of the operator 'N' under which lines 4 and 6 are both true and rule (β) is also valid – just as he claims that there is no single construal of 'can render false' under which Premise 5 and Premise 6 of the earlier argument are both true.

NOTES

[1] Van Inwagen, Peter: 1975, 'The incompatibility of free will and determinism', Philosophical Studies 27, pp. 185–199; van Inwagen, Peter: 1983, An Essay on Free Will (Oxford), henceforth EFW.

[2] Lewis, David: 1981, 'Are we free to break the laws?', Theoria 3, pp. 113–121.

[3] Ginet, Carl: 1966, 'Might we have no choice?', in K. Lehrer (ed.): Freedom and Determinism (Random House, New York), pp. 87–104; Wiggins, David: 1973, 'Towards a reasonable libertarianism', in T. Honderich (ed.): Essays on Freedom of Action (Routledge and Kegan Paul, Boston), pp. 31–62; Lamb, J. W.: 1977, 'On a proof of incompatibilism', Philosophical Review 86, pp. 20–35; Ginet, Carl: 1980, 'The conditional analysis of freedom' in Peter van Inwagen (ed.): Time and Cause (Reidel, Dordrecht), pp. 171–186.

[4] Slote, Michael: 1982, 'Selective necessity and the free-will problem', Journal of Philosophy, pp. 5–24.

[5] In this version of the argument, which appears in his book, Premise 2 is somewhat

stronger than it was in the original article. I explain the reason for this change in Note 12 below.

[6] What he actually says, on p. 192 of van Inwagen 1975, is that the following general principle is analytic:

> If p is a true proposition that concerns only states of affairs that obtained before S's birth, and if S can render the conjunction of p and q false, then S can render q false.

But it is hard to see why someone would consider this principle analytic if he did not also consider both (A) and (B) analytic.

[7] Slote, 1982, points out that every version of the Consequence Argument he knows of employs some version of the following modal principle, which he calls *agglomerativity*: If Nec(p) and Nec(q), then Nec(p and q). Principle (A) is just such an agglomerativity principle. This becomes clear when we state it in the equivalent contraposed form: If S cannot render p false, and S cannot render q false, then S cannot render (p and q) false. Thus, the fact that agglomerativity does not hold, under Lewis's definition of 'can render false in the strong sense', is highly relevant to versions of the Consequence Argument other than the one we are examining here.

[8] Lewis defends this view in Lewis, David: 1979, 'Counterfactual dependence and time's arrow', Nous 13, pp. 455–476.

[9] See Bennett, Jonathan: 1984, 'Counterfactuals and temporal direction', Philosophical Review 93, pp. 57–91. In this paper Bennett replies explicitly to Lewis's arguments in Lewis, 1979.

[10] John Martin Fisher replies in much this way to the version of the Consequence Argument given in Ginet, 1980. See Fischer, John Martin: 1983, 'Incompatibilism', Philosophical Studies 43, pp. 127–137. His reply rests on a distinction very much like Lewis's distinction between the weak and strong senses of 'can render false'; see p. 130.

[11] Van Inwagen, Peter: 1977, 'Reply to Narveson', Philosophical Studies 31, pp. 89–98.

[12] Van Inwagen, 1977, p. 93. This definition is what necessitates the strenthening of his original Premise 2, which was 'If J had raised his hand at T, then P would be false'.

[13] I take it that there is no important difference, at least none that matters for our purposes here, between what Lewis means by 'event' and what van Inwagen means by 'state of affairs'.

[14] If one adopts the nonmiraculous analysis of counterfactuals, then van Inwagen's official definition of 'can render false' fails to meet even minimal standards of intuitive plausibility, and therefore cannot serve his purposes. Under his definition, altered-past compatibilists wind up committed to the proposition that if determinism is true then J could have rendered L false. (This is because it not possible that J should raise his hand at T, and the past be the same as it actually was, and L be true.) They are committed to this proposition even though they vigorously deny that L would have been false if J had raised his hand at T. In the face of this result, they can surely claim that van Inwagen's definition is just too Pickwickian to be plausible or interesting. Accordingly, they can deny Premise 6 relative to van Inwagen's definition, even though they are prepared to accept Premise 6 relative to any definition they consider reasonable. (Under a reasonable definition, a necessary condition for the truth of the proposition expressed by 'J can render L false' is that J can do something such that, if he did it, L would indeed be false.)

If one adopts the miraculous analysis, on the other hand, then van Inwagen's definition of 'can render false' is more plausible: for, under this analysis his definition is considerably closer to the definition of 'can render false in the strong and broad sense' which I am about to propose. But his definition is still somewhat inadequate intuitively: for instance, it commits the divergence-miracle compatibilist to saying that if determinism is true then J can render false the proposition that J's acts are never law-violations – where a law-violation is an event which falsifies L in Lewis's strict sense. (This is because

necessarily, if J raises his hand at T and the past is exactly as it was in the actual world (so that there is no prior divergence miracle), then J's raising his hand at T is itself a law-violation.) In any case, ultimately it doesn't much matter, from the perspective of divergence-miracle compatibilism, whether van Inwagen's definition is intuitively reasonable or not. For, I shall argue below that if the miraculous analysis of counterfactuals is correct, then Premise 6 is false under the definition of 'can render false in the strong and broad sense' I shall now propose. And a parallel argument can be used to show that Premise 6 is also false under van Inwagen's official definition.

[15] Some of the claims in this and the next paragraph only hold if determinism is true. But what we are interested in knowing, of course, is what the compatibilist will say about Premise 5 and Premise 6 when he is supposing that determinism is true.

[16] After this paper had gone to press I sent a copy to David Lewis, who sent me a set of detailed, and very helpful, comments. I thank him for them, and I shall briefly remark upon them here. In relation to the Nostradamus example, he points out that Nasser can indeed render N_2 false in the weak sense; it's just that the falsifying-event will be not merely the destruction of the Sphynx, but rather a complex event which includes the destruction and also includes a subsequent chunk of human history. This is true, and I was mistaken to suggest otherwise. Still, it seems initially that there should be a *strong* sense in which Nasser can render N_2 false – a sense in which the N_2-falsifying event is actually an effect of Nasser's own act. And the Martian example does seem to show that Lewis's own definition of 'can render false in the strong sense' does not fit the bill. This suggests that there might be a sense of 'can render false' which is weaker than Lewis's strong sense, stronger than Lewis's weak sense, and capable of rendering all of van Inwagen's premises simultaneously plausible. The appropriate candidate, I suggested, was my 'can render false in the strong and broad sense': but it turned out not to serve van Inwagen's purposes after all.

Lewis also mentions a very weak sense of 'can render false': an agent can render a proposition false in the *simple* sense iff he is able to do something such that, if he did it, the proposition would be false. He demonstrates that under some plausible assumptions, my "strong and broad sense" of 'can render false' (or rather, a slight revision of it which he argues is independently motivated) turns out to be equivalent to the simple sense: and so does my "weak and broad sense." I take it that this (somewhat surprising) result actually reinforces the point I make at the end of the paper – *viz.*, that definitional maneuvering cannot help van Inwagen evade the sort of objection which Lewis originally raised. It also suggests that, contrary to one's initial expectations, there probably is no interesting sense of 'can render false' which is stronger than Lewis's weak sense but weaker than his strong sense.

[17] In this paragraph I have used 'P_0', 'L', and 'P' in each of van Inwagen's two ways. I trust that my usage is clear in any given instance.

[18] This fact is pointed out by Slote, 1982, who stresses that most versions of the Consequence Argument employ a modal principle like (β). He gives examples of various modalities which allegedly do not obey this principle, and then provides reasons for thinking that modalities like van Inwagen's 'N' do not obey it either. His discussion is similar in spirit to Lewis, 1981 and to the present paper, although he does not say whether it is agglomerativity or closure under entailment which should be rejected, and he does not consider the possibility of interpreting modalities like 'N' in such a way that (β) is valid but either 'NP_0' or 'NL' is false.

Department of Philosophy,
Memphis State University,
Memphis, TN 38152,
U.S.A.