

REPLY TO NARVESON

(Received 22 November, 1976)

This paper is in two parts. In Part I, I discuss two important points on which Jan Narveson and I disagree.<sup>1</sup> In Part II, I discuss what I think is the *central* point on which we disagree.

I

(i) *I ... want to take issue with van Inwagen's claim that his argument, if valid, would still not lead to the conclusion that there are 'uncaused events'.* (p. 85) I think Narveson wants to take issue with my claim (IFD, p. 196) that the denial of determinism does not entail that there are uncaused events. (This claim, if true, blocks the inference from 'Van Inwagen's argument is sound' and 'We have free will' to 'There are uncaused events'.) But if this *is* what he wants to take issue with, it is puzzling that he should ask, "Does he mean that in the ordinary sense of 'cause' ... his definition [of determinism] does not imply the definition of determinism in terms of every event having a cause?" (p. 86) This question seems to be equivalent to the question whether I think *this* is true: Determinism (in the sense defined in IFD) does not entail universal causation. An interesting question, but irrelevant to my argument. The relevant question is whether I think *this* is true: The denial of determinism does not entail the denial of universal causation. And the answer is, that's what I *said* and that's what I think.

Why do I think this? Because to infer the denial of universal causation from the denial of determinism, one needs at least three additional premises:

- P(1) If an event (or fact, change, state of affairs, or what have you) has a cause, then its cause is always itself an event (or what have you) and never a substance, such as a man
- P(2) If event (or what have you) *A* was the cause of event *B*, then it follows, given that *A* happened and given the laws of physics, that *B* could not have failed to happen, that *A* 'causally necessitated' *B*

- P(3) Every chain of causes that has no earliest member is such that, for every time  $t$ , some event in that chain happens earlier than  $t$

and I doubt whether *all three* of these premises are true. I am in particular doubt about P(2). I shall not here explain why these three premises are required for the deduction of ‘Some event has no cause’ from ‘Determinism is false’ or defend my skepticism about them. I refer the interested reader to R. M. Chisholm, ‘Freedom and Action’<sup>2</sup> [P(1)], G. E. M. Anscombe, *Causality and Determination*<sup>3</sup> [P(2)], and Jan Łukasiewicz, ‘On Determinism’<sup>4</sup> [P(3)].

An incidental note: in his argument of p. 86, Narveson passes from saying that if my arguments are sound and if we have free will then “some events are *inexplicable* in terms of antecedent events and physical laws”, to suggesting that, if these two conditions are satisfied, then some events are inexplicable *simpliciter*. And he gives no argument to show that such a move is legitimate. There are other sorts of explanation than nomological-deductive explanation. Does Narveson want to say that if no nomological-deductive explanation can be given for some event, it follows that no teleological (say) explanation can be given for that event? I should like to see an argument for that.

(ii) Narveson objects to my claim that, while one is necessarily unable to render false any proposition that is a physical law, there might well be psychological laws that one *could* render false (IFD, p. 187).

*After all, what is there about the ‘laws of physics’ that is relevant to his argument? Only one thing: viz., that given the set of them as assumed premises, one can deduce any proposition about the state of the universe at a particular time from any proposition about the state of the universe at any other time.* (p. 86) The feature Narveson attributes in this passage to the set of physical laws is not a necessary or even an actual property of this set. It is a feature of the laws of physics only in those possible worlds in which determinism is true, a category into which, if current physics in anything other than a ghastly illusion, the actual world does not fall. A property of physical laws that is rather more obviously ‘relevant’ to my argument than the one Narveson mistakenly attributes to them is this: No one can render any of them false. And the laws of physics have this property (I claim) of conceptual necessity: they have it in all possible worlds, including indeterministic worlds like the actual world. Consider an indeterministic set of physical laws that en-

tails that if *A* happens, either *B* or *C* happens, and which entails *neither* that if *A* happens, *B* happens *nor* that if *A* happens, *C* happens. Nothing in the supposition that these laws are the *only* physical laws renders the following impossible: Given that *A* happens, it is up to us, we have a choice, whether *B* happens or *C* happens. But our supposition does render *this* impossible: Given that *A* happens, it is up to us whether either *B* or *C*, on the one hand, or *something else*, on the other, happens. And this limitation on what is up to us, on what we have a choice about, is *not* due to its being the case that 'If *A* happens, then *B* or *C* happens' is a consequence of a *deterministic* set of physical laws (for we have supposed that the set of physical laws is *not* deterministic); it is due to the fact that this proposition is a consequence of a set of *physical* laws.

*Insinuating that 'voluntaristic laws' would have the curious property of being changeable at the will of the agent is, after all, simply admitting that they are not laws: if I can do something which is contrary to a general statement about my behavior, I thereby show that this statement is false, i.e., that it is not a law of my behavior.* (p. 86) A key clause in this passage ("if I can do something ... statement is false") resists my best efforts to interpret it. (Cf. 'If I can insult Jones, I thereby offend him'.) Perhaps Narveson means only: 'If I *do* something which is contrary to a general statement about my behavior, I thereby show that this statement is false'. This is true, and I won't deny it. But I *will* deny the following proposition which I think Narveson would want to defend: If I *can* (but don't) do something contrary to some (true) general statement about human behavior, then it follows that this statement is not a law.

Suppose psychologists discover that anyone who received moral training of type *A* in early childhood never spreads lying rumors about his professional colleagues. Suppose you and I in fact received such training. Does it follow that we *can't* engage in this shoddy activity? I don't see why it should be supposed to follow. (Mark Twain: "I am morally superior to George Washington. He couldn't tell a lie. I can and I don't.") Suppose further that you and I are in fact *able* to spread lying rumors about our colleagues. Does it follow that a statement of the regularity we imagined psychologists to have discovered is (though true) not a law? Well, suppose the existence of this regularity is a logical consequence of some generally well-confirmed theory of human moral-development that has great explanatory and predictive power. Why *not* call a statement of it a law? "But why", someone may ask, "does this regularity

occur in human behavior if people don't *have* to conform to it?" Note that the only people in a position to depart from this regular pattern of behavior are those who have in fact had training of type *A*. Perhaps it is just these people who *see the point* in not spreading lying rumors. To come to see the point in not exercising a certain ability one has is not to *lose* that ability.

## II

Let us now turn to what was called in IFD the 'Main Argument' and to my defense of its premises. Narveson thinks this argument is objectionable, but I am not confident I understand his criticisms of it. In the sequel, I shall defend my argument against an objection that is at least suggested by what Narveson says on p. 85. Before I do this, however, it will be necessary to get clear about how the Main Argument works. I had thought this was clear, but Narveson, in his description of my paper, says things that show that it is not clear to him.

*He then presents [a] seven-step argument, the seventh of which is alleged to be a consequence of the first six on any relevant analysis of 'can' ... (p. 83)* Since the argument Narveson refers to (the Main Argument) is formally valid, this is like saying that 'God exists' is 'alleged to be' a consequence of 'If there is motion, God exists' and 'There is motion' on any relevant analysis of 'motion'. I *do* claim that the six premises of the argument are *true* on any (correct) analysis of 'can' in the relevant *sense* (the 'power' or 'ability' sense).

*But what about (4)? Does it follow from (3)? (p. 84)* Of course not. I never claimed it did. On pp. 191–194 of IFD, I defended (1)–(6) severally. I don't claim that any of them follows from any combination of the others. Nor do I claim that (7) follows from any proper subset of (1)–(6), nor that any principles other than those of sentential logic are needed for the deduction of (7) from (1)–(6). Narveson calls the premises of the Main Argument 'steps'. This gives me the impression that he may have failed to appreciate these facts. This impression is reinforced by what he says on p. 85 about the relations between (5), (6), and (7).

So the only important question is this: Is any of (1)–(6) *false*? Narveson thinks that premise (4) of the Main Argument is false. More exactly, he thinks (4) is false if 'render' is understood in its usual sense; and that, if 'render' is understood in some *outré* sense, so as to make (4) analytic, then (5) is false. Perhaps I did not explain my use of this word with sufficient care. I certainly gave no explicit definition of 'could have rendered false'. Nevertheless, what

I said in my discussion of this phrase in IFD (see pp. 189–190 and note 5) makes it tolerably clear that I meant ‘*S* can render *p* false’ to mean ‘Either *p* is false, or else there is some set of conditions *C* such that (a) *S* can produce *C*, and (b) *C* is sufficient for the falsity of *p*’.

But perhaps this definition is a bit too vague for our present purposes. In particular, ‘sufficient’ is vague. Let us tighten up this definition.<sup>5</sup> Let us first say that a state of affairs *entails the falsity of* a proposition *p* just in the case that it is not possible (in, to borrow a phrase of Plantinga’s ‘the broadly logical sense’) that that state of affairs obtain *and p* be true. Now let us say that sentences of the form ‘*S* can render *p* false’ are to be regarded as abbreviations for the corresponding sentences of the form

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*.<sup>6</sup>

This definition is intended to be an adequate vehicle for the expression of the following intuitive considerations. Consider the propositions I should express if I were to utter any of the following sentences at the present moment

- (a)  $27 \times 15 = 405$
- (b) Magnets attract iron
- (c) Mary, Queen of Scots was put to death in 1587
- (d) I have never read *The Teachings of Don Juan*
- (e) No one has ever read all of Hume’s *Enquiry* aloud
- (f) The cup on my desk has never been broken.

(All these propositions are, I think, true.) There seems to me to be at least one important and interesting difference between the relations I bear to (a)–(c) and those I bear to (d)–(f). The difference I have in mind might be put this way: There is nothing I can do, or could ever have done, about the fact that (a)–(c) are true, and this is *not* the case with (d)–(f). (In saying this, I assume I have free will. If I don’t, then, of course, this apparent difference is illusory.) The ‘can-render-false’ idiom of IFD was chosen as a device for expressing this distinction: one may express it in that idiom by saying, “While van Inwagen can render, or at least, once could have rendered, (d)–(f) false, he cannot render, and never could have rendered, (a)–(c) false.” Moreover, I believe that the above paraphrase of ‘*S* can render *p* false’ succeeds in

marking out just the relation between agent and proposition that I am interested in. (I want to define a relation that holds between agents and *propositions* because I am investigating the conceptual relationship of free will, a thesis about agents, to determinism, a thesis about propositions.)

Now it may be that I have chosen a misleading *name* for this relation. Well, anyone who does not like my name for it is free to call it ‘Charley’. After all, the ‘can-render-false’ idiom does not appear in the *conclusion* [(7)] of the Main Argument; it has ‘dropped out’, and only ordinary words meant to be taken in their ordinary senses are left. The only *important* questions about ‘*S* could have rendered *p* false’ are ‘Is the paraphrase of this sentence-form sufficiently clear that the premises of the Main Argument that contain clauses of that form have truth-values?’ and ‘Given this paraphrase, are they in fact true?’ I think the answer to these questions is Yes.

Narveson, as I said, objects to premise (4). This premise may be defended as follows (cf. the defense of (4) in IFD, p. 192). Premise (4) is an instance (allowing for a shift of tense) of the following general principle:

- (g) If *S* can render *R* false, and if *Q* entails *R*, then *S* can render *Q* false.

(To see this, substitute ‘*J*’ for ‘*S*’, ‘*P*’ for ‘*R*’, and ‘the conjunction of *P*<sub>0</sub> and *L*’ for ‘*Q*’.) Principle (g) is clearly analytic. For if *Q* entails *R*, then the denial of *R* entails the denial of *Q*. Therefore, if there is some state of affairs *S* can bring about that entails the falsity of *R*, there is some state of affairs (that same state of affairs) that *S* can bring about that entails the falsity of *Q*.

So (4) follows from the analytic principle (g). Narveson thinks that if I ‘insist’ that (g) is analytic, this insistence will have untoward consequences for premise (5) of the Main Argument. I am not sure I understand his reasoning (p. 85), but perhaps it could be represented as follows.<sup>7</sup> “Very well, you have stipulated a sense for ‘can render false’ such that (g) is, if interpreted according to your definition, analytic. But then the following principle

- (h) If *Q* is a true proposition that concerns only states of affairs that obtained before *S*’s birth, and if *S* can render the conjunction of *Q* and *R* false, then *S* can render *R* false [IFD, p. 192],

the analyticity of which is essential to your defense of your premise (5), is certainly not analytic. For (h) is analytic only if

- (i) If  $Q$  is a true proposition that concerns only states of affairs that obtained before  $S$ 's birth, then  $S$  cannot render  $Q$  false

is analytic, which it is not, at least if it is (as it is supposed to be) short for

- (i\*) If  $Q$  is a true proposition that concerns only states of affairs that obtained before  $S$ 's birth, then there is no state of affairs  $A$  such that (a)  $S$  can bring about  $A$ , and (b)  $A$  entails the falsity of  $Q$ .

For if determinism is true (and determinism is at least *possibly* true) then the universal closure of (i\*) is false. Consider some state of affairs  $H$  that obtains just in the case that my hand is raised one minute from now, which it will not be. Let  $\mathcal{L}$  be the state of affairs that consists in the laws of nature being just what they in fact are. Let ' $H \& \mathcal{L}$ ' designate the conjunction of, or joint obtaining of, these two states of affairs. Let  $P_1$  be some proposition that expresses the state of the world before my birth. Then I can render  $P_1$  false, contrary to the claim made by (i). For there is a state of affairs —  $H \& \mathcal{L}$  — such that I can bring this state of affairs about, and such that this state of affairs entails the falsity of  $P_1$ . I can bring about  $H \& \mathcal{L}$  because I can raise my hand one minute from now, and if I were to raise my hand one minute from now,  $H \& \mathcal{L}$  would obtain. I can raise my hand one minute from now because *if* I were to choose then to raise my hand, I should then raise it, and this is what it *means* to say I can raise my hand one minute from now."

To put this objection to the Main Argument in a nutshell: Given my definition of 'can render false' and given determinism, one can render at least some propositions about the past false, and if this seems counterintuitive, that the fault of my definition of 'can render false'.

If the reasoning I have attributed to Narveson is correct then what follows is that *if* the 'conditional analysis' of ability is correct, *then* premise (5) of the Main Argument is false. (Or, if (5) is not false, it is at least the case that principle (h), the analyticity of which was essential to my *defense* of (5), is not analytic.) Now in IFD (p. 197), I conceded that it might very well be the case that the conditional analysis entails the falsity of some premise of the Main Argument. I said that if this should turn out to be the case, then I should simply regard the 'falsified' premise as a counterexample to the conditional analysis.

I have not changed my mind about this. I continue to regard (5) as true and principle (h) as analytic. Therefore, it seems to me, *if* the conditional

analysis is in conflict with (5) or with (h), then the conditional analysis is wrong. I say this because (h) seems to me to be more plausible than the conditional analysis.

What I have said in the last two paragraphs is merely an application to (5) and (h) of some very general remarks I made at the close of IFD. Let us consider the matter in greater detail.

Principle (h) is, I should think, analytic if and only if (i) is. And (i) is analyzed as (i\*). And (i\*) seems to me to be obviously analytic. If there is some state of affairs that entails the falsity of some *true* proposition about the way the world was before I was born, then I can't bring about (and never could have brought about) this state of affairs. If the preceding sentence expresses a falsehood if 'I can' means 'I should if I chose', then 'I can' does *not* mean 'I should if I chose'.

Or put the matter this way. Here are two theses, formally about the English language, but not, perhaps, devoid of metaphysical implications:

The English sentence 'If some state of affairs entails the falsity of some true proposition about the way the world was before I was born, then I can't bring about that state of affairs' is analytic.

The English words 'I can' mean the same as 'I should if I chose'.

The former of these theses seems to me to be obviously true (at least in so far as 'analytic' has a clear sense). The latter I have no particular intuitions about. Therefore, *if* these two theses are indeed inconsistent, I unhesitatingly choose the former. I have no idea how to argue for these intuitions in any way that is both clear and non-question-begging. I *could* say that the past is 'immutable' or some such, but this would not be very helpful, since it's not clear what it means. Or I could say, "Obviously, if some state of affairs is (in 'the broadly logical sense') sufficient for Mary, Queen of Scots' having died a natural death, then it is not within my power to bring about that state of affairs", but this, perhaps, would be to beg the question. I think I can do no more than to examine carefully any arguments that may be offered for the conclusion that my intuitions about abilities and the past are incorrect. I have not yet seen any.

One other matter. It is not at all clear that it follows from the truth of determinism and the correctness of the conditional analysis that I can render false some proposition about the past. Let us return to the example about my

raising my hand that figured in my representation of Narveson's argument *contra* (i\*). An essential premise of this argument was the contention that, if I can raise my hand one minute from now, then I can bring about  $H \& \mathcal{L}$ . This seems to me to be quite doubtful given that determinism is true and  $H$  does not obtain. What seems to me to be more likely to be the truth is this: if I can raise my hand one minute from now, then I can bring it about that  $\mathcal{L}$  does not obtain; that is to say, if I can raise my hand one minute from now, then I can render  $L$  false.

Let me give an example in support of this intuition. The obtaining of  $H$  is 'ruled out' by the state of the world at some past moment, taken together with the laws of physics, only assuming that the laws of physics are deterministic. Let us look at a case in which a certain physically possible<sup>8</sup> future state of affairs is 'ruled out' by the actual state of the world one minute ago taken together with a certain proposition commonly held to be a law of physics. Consider the state of affairs that consists in my touching the Viking I lander on Mars one minute from now. That this state of affairs does not obtain is entailed by the proposition that nothing travels faster than 186 000 miles per second (commonly held to be a law of physics) taken together with the state of the world one minute ago (one feature of the world one minute ago was that the Viking lander and I were at that moment separated by a distance greater than 22 320 000 [=  $2 \times 60 \times 186\,000$ ] miles). Well, of course, I *can't* touch the Viking lander one minute from now. But suppose for the sake of argument I *can*. What follows from this supposition? It seems to me to follow that I *can* travel faster than 186 000 miles per second. Let us put this conclusion in terms of states of affairs, in order to underline the formal analogies between this case and the  $H$ -and- $\mathcal{L}$  case. Let  $M$  be the state of affairs that consists in my touching the Viking lander one minute from now. Let  $\mathcal{L}^*$  be the state of affairs that consists in its being the case that nothing travels faster than 186 000 miles per second. Suppose for the sake of argument I can bring about  $M$ . Does it follow that I can bring about  $M \& \mathcal{L}^*$ ? It hardly *seems* to. What seems to follow is that I can bring it about that  $\mathcal{L}^*$  does not obtain.

To return to the general case, what seems to me to emerge is this: if determinism is true and if the conditional analysis is correct, then what follows is *not* that I *can* (though, of course, I *shall* not) bring about some state of affairs that entails the falsity of some true proposition about the past. Rather, what follows is that I *can* (though, of course I *shall* not) bring

about some state of affairs that entails the falsity of some law of physics. Now since I don't think I can (even if determinism is true) bring about any state of affairs that entails the falsity of some physical law, I think the conditional analysis must be wrong.

*Syracuse University*

#### NOTES

<sup>1</sup> See Jan Narveson, 'Taking Compatibilism Seriously', *Philosophical Studies*, this issue. Narveson's article is a criticism of my 'The Incompatibility of Free Will and Determinism', *Philosophical Studies* 27 (1975) pp. 185–199, hereinafter cited as IFD. In the sequel, I presume familiarity with the terminology and conventions of IFD. Propositions referred to by number (e.g., '(5)') are those so numbered in IFD.

Lengthy passages in italics are quotations from Narveson's paper.

<sup>2</sup> In K. Lehrer (ed.) *Freedom and Determinism* (New York, 1966).

<sup>3</sup> Inaugural lecture, published as a pamphlet by Cambridge University Press (1971).

<sup>4</sup> In Storrs McCall (ed.) *Polish Logic 1920–1939* (Oxford, 1967). See especially pp. 30–32. For a good recent discussion of this point, see Jordan Howard Sobel, 'Determinism: A Small Point', *Dialogue* XIV (1975) pp. 617–621.

<sup>5</sup> This tightening requires a minor revision of premise (2) of the Main Argument: this premise, which is a subjunctive conditional, should be replaced with the corresponding strict conditional.

<sup>6</sup> In the language of my paper, 'A Formal Approach to the Problem of Free Will and Determinism', *Theoria* (1974, Part I) pp. 9–22, '*S* can render *p* false' may be paraphrased as ' $\exists x(H(Sx) \ \& \ p \text{ is false in } x)$ ', if we ignore some problems about tenses. All the verbs in the paraphrase given in the present paper are to be regarded as tenseless, except for 'can', which is, of course, in the present tense. A paraphrase of '*S* could have rendered *p* false' can be got by replacing 'can bring about' in the paraphrase in the text with 'could have brought about'.

I have given an explanation of sentences of the form '*S* can render [or *could have rendered*] *p* false', and this does not provide us with any obvious way of understanding sentences of the form '*S* does render *p* false'. But this is not a very serious omission, since nowhere in the Main Argument or in my defense of its premises does there occur any sentence of the latter form.

<sup>7</sup> If what follows is not a representation of Narveson's argument, then I do not know what his argument is. At any rate, it is an interesting argument and I welcome the chance to say something about it. The possibility of arguing in this way was pointed out to me in conversation by Raymond Martin and Michael Gardner several years ago.

<sup>8</sup> This future state of affairs is 'physically possible' in the sense that there are possible worlds in which it obtains in which the laws of physics are the same as those of the actual world. There is, however, another sense of 'physically possible' in which it is *not* physically possible. See note 3 (p. 12) to the article cited in note 6 above.