Δαίμων 3 – 1991. Págs. 57-63

Modal Inference ant the Free-Will Problem

PETER VAN INWAGEN

ABSTRAT

It has long been held that arguments for the incompatibility of free will and determinism depend upon a modal fallacy. This charge was revived by Michael Slote in his important 1982 article, «Selective Necessity and the Free-Will Problem». In that article, Slote criticized my own work and the work of several other philosophers who had argued for the incompatibility of free will and determinism. In the present article, I defend my arguments against Slote's charge that they involve invalid modal reasoning.

I shall begin by laying out what seems to me to be a good, though hardly indisputable, argument for the incompatibility of free will and determinism [1].

Let 'N' be an operator that expresses whatever sort of necessity it is that is opposed to free will. It seems plausible to suppose that the following two inference-rules governing 'N' are valid:

(α) $\Box p, \vdash Np$ (where '□' has its standard sense)
(β) $Np, N(p \supset q) \vdash Nq$.

Now let 'S' be a sentence that gives a complete description of the state of the world at some time in the remote past. Let 'L' be a sentence that expresses the conjunction into a single proposition of the laws of nature. Let 'T' be any sentence that expresses a truth about the present or the future. Now assume determinism. It is a consequence of

1 This argument is discussed in greater detail than is possible (or necessary) here in Chapter III of my book *An Essay on Free Will* (Oxford: The Clarendon Press, 1983).

determinism that

$$\Box ((S \cdot L) \supset T).$$

And from this consequence of determinism, it follows, by elementary modal and sentential logic, that

$$\Box (S \supset (L \supset T)).$$

From this it follows by (α) that

$$N(S \supset (L \supset T)).$$

Now it would seem that both the laws of nature and statements about the past – and certainly the remote past – should be accounte «necessary» in the sense expressed by 'N' [2]. We have, therefore 'NS' and 'NL', and then, by two applications of (β), 'NT'. That is, if determinism is true, then any truth whatever is «necessary» in a sense that is opposed to our having free will about whether it is true. If the argument that has led to this conditional is valid, then determinism is incompatible with free will.

Is the argument valid? Since no one (I think) would want to dispute (α), this question reduces to the question whether (β) is valid. And that is a very good question indeed.

I

In an important and much-cited article, Michael Slote has attempted to cast doubt on the validity of (β) [3]. Slote suggests that anyone who accepts (β) probably accepts it only because he accepts these two rules:

---

2 «Not even the gods can change the past»; *only* a God – or, at least, some sort of supernatural being – could change (or violate, set aside, defy, cause things to act in ways other than those prescribed by) the laws of nature.

3 «Selective Necessity and the Free-Will Problem», *The Journal of Philosophy*, LXXIX, 1 (January 1982) pp. 5-24. Slote's article is a criticism of four defenses of the incompatibility of free will and determinism: Carl Ginet, «Might We Have No Choice?» in Keith Lehrer, ed., *Freedom and Determinism* (New York: Random House, 1966), pp. 87-104; James Lamb, «On a Proof of Incompatibilim», *Philosophical Review*, LXXXVI, 1 (January 1977), pp. 20-35; David Wiggins, «Towards a Reasonable Libertarianism», in Ted Honderich, ed., *Essays on Freedom of Action* (London: Routledge & Kegan Paul, 1973), pp. 31-61; my «The Incompatibility of Free Will and Determinism», *Philosophical Studies*, XXVII, (March 1975) pp. 185-199. My article is reprinted in Gary Watson, ed., *Free Will* (New York: Oxford, 1982), pp. 46-58.

Slote has a certain amount of trouble with my article, the argument of which is laid out in a way that is rather inconvenient for his purposes. But he is perfectly right in thinking that the points he makes are as

Agglomeration $Np, Nq \vdash N(p \cdot q)$

Closure $Np \vdash Nq$, provided q is derivable from p.

He says:

> Anyone who assumes the validity of arguing from 'Np' and '$N(p \supset q)$' to 'Nq' would seem to be tacitly assuming that the necessity expressed in the operator 'N' is both agglomerative (closed with respect to conjunction introduction) and closed under logical implication, so that one can, e. g., validly move from 'Np' and '$N(p \supset q)$' to '$N(p \cdot p \supset q)$' and from the letter to 'Nq'. If we do not think about these subinferences, when we move from 'Np' and '$N(p \supset q)$' to 'Nq' or assert the... modal principle [(β)] that corresponds to that larger inference, that is only because it is so natural to assume that any necessity operator will have the properties of agglomerativity and closure under logical implication or entailment.

If I understand my own allegiance to (β), however, it does not seem to have come about in this way. Let me try to explain this allegiance. To begin with, I believe that the necessity that (as I rather vaguely put it above) «is opposed to free will» should be spelled out in this way:

$Np =$ df p and no one has, or ever had, any choice about whether p.

At any rate, an argument for the conclusion that determinism entails that every true proposition is necessary in this sense is certainly correctly describable as an argument for the incompatibility of free will and determinism. (Moreover, the premises of the argument for the incompatibility of free will and determinism that I laid out above —'NS' and 'NL' – would seem to be obviously true on this interpretation of 'N', at least if we interpret 'S' as a statement about the state of the world before there were any beings of the sort whose free will we are interested in.)

I believe that (β) is valid if the operator 'N' is interpreted in this way. (Let us say that on this interpretation, 'N' expresses *Choice Necessity*). But, so far as I can tell, this conviction of mine does not arise from a prior conviction, or even a tacit assumption, that Agglomeration and Closure, or any other inference-rules, are valid [4]. I believe that

---

applicable to my article as the other three. Because the argument of Section 3.10 of my book (see n.º 1) is laid out in such a way that Slote's points can be applied to it without trivial and annoying adjustments and qualifications, I shall defend that argument rather than the argument of «The Incompatibility of Free Will and Determinism».

4 Nor, as Slote repeatedly suggests, does the example of standard alethic modal logic play any role in my conviction. Or not so far as I can tell.

this conviction arises from the intrinsic plausibility of (β) when 'N' is interpreted as expressing Choice Necessity. (In the sequel, 'N' is to be understood as expressing Choice Necessity – and (β) is to be interpreted accordingly – unless I explicitly stipulate some other interpretation.) I cannot, in the strictest sense of 'argument', give an argument for the thesis that (β) is valid, for I know of no thesis more plausible than itself from which it follows. But I may be able to communicate my conviction that (β) is valid, and communicate it in a very simple and straightforward way. If you wish to appreciate the plausibility of the validity of (β), attempt to construct a counterexample to it. I believe that the reader who has made a serious and sustained effort to construct a counterexample to (β) will come to share my conviction that (β) is valid; or, if not to share it, then at least to see how someone could have this conviction quite independently of his convictions about the validity of any other rules of inference [5].

I will remark in passing that I *am* inclined to think that Agglomeration and Closure are valid. Consider the set of worlds W such that the actual world belongs to W and a non-actual world belongs to W if and only if someone has, or once had, a choice about whether that world is actual. It seems to me to be plausible to suppose that W is such that, for any *p*,

p and no one has, or ever had, any choice about whether *p* is true

(*sc.* in the actual world) if and only if *p* is true in every member of W [6]. That is to say, there is a certain set of worlds (containing the actual world) such that for every *p*, 'N*p*' is true in every member of that set. More generally, for every world *w*, there is a set of worlds *W* (containing *w*) such that for every *p*, 'N*p*' is true in *w* if and only if *p* is true in every member of *W*. This means that 'N' is what we might call a «classical» neces-

5 I am not offering an inductive argument. I am not proposing that the reader, having made (say) sixteen failed attempts to construct a counterexample to (β), should reason as follows: «All sixteen of my attempts to find a counterexample to (β) have been failures. Therefore, probably, (β) is valid». Rather, my hope is that the reader, in the course of attempting to construct counterexamples to (β), will come to *see* why (β) is valid.

Here is a general description of what one experiences when one attempts to construct a counterexample to (β). One substitutes particular declarative sentences for '*p*' and '*q*', and one devises a case according to which, one hopes, the premises of the resulting argument are true and its conclusion false. On careful examination, however, it transpires that (say) one of the premises is false; one adjusts the case to make the premise true and thereupon discovers that one has inadvertently made the other premise false; a second adjustment, intended to correct this fault, causes the conclusion to be true; when *this* defect is corrected, it turns out that the premise that occasioned the original modification of the case is once more false. As the hours pass – hours during which one tries many and various pairs of substitutions for '*p*' and '*q*' and constructs many bizarre scenarios –, one begins to recognize patterns in the repeated blockings of the «TTF» case, patterns that display to one the inevitability of the frustration of every attempt to devise an instance of that case.

6 I shall not invariably be careful about distinguishing use and mention or variables from dummy letters.

sity operator and not what Slote calls a «selective» necessity operator. I leave as an exercise the trivial proof that Agglomeration and Closure hold for all classical necessity operators.

I am therefore inclined to think that 'N' is, as Slote puts it, «both agglomerative and closed under logical implication» because I am inclined to think that the following two operators are equivalent:

p and no one has, or ever had, any choice about whether *p*.

The proposition that *p* is true in the actual world and in all non-actual worlds such that someone has, or once had, a choice about whether they are actual.

But if someone could convince me that this equivalence did *not* hold, and could, in fact, convince me of the truth of the stronger statement that the Choice Necessity operator was not a classical necessity operator, I should still accept (β). (We should note that while the statement that 'N' is a classical necessity operator entails the validity of (β), the validity of (β) does not entail that 'N' is a classical necessity operator [7]. I should still accept (β) because my conviction that (β) is valid rests on what I believe I have learned by attempting to construct counterexamples to (β) and not on my belief that the two sentences displayed above are equivalent or on my belief that the Choice Necessity operator is a classical necessity operator. (My belief that there are people does not rest on my belief – or on my tacit assumption – that there are people in Tibet, despite the fact that I do believe, and with great conviction, that there are people in Tibet, and know that the proposition that there are people in Tibet entails the proposition that there are people. If something convinced me that there were, after all, no people in Tibet, I should still believe that there were people.)

I have implied that it is at least very difficult to find a counterexample to (β). I will now say this explicitly, and while I am at it, make explicit an important qualification: it is at least very difficult to find a counterexample to (β) that can be seen to be a counterexample independently of the question whether free will is compatible with determinism. Of course if free will is compatible with determinism, it is easy to find a counterexample to (β); in fact we have already done so. If free will is compatible with determinism, then one or the other of the two aplications of (β) in the argument for incompatibilism displayed above must have taken us from truth to falsity. But, of course, putative counterexamples to (β) that «work» only if compatibilism is true are of no interest in a dispute about the truth of compatibilism. What would be of interest would be a putative counterexample to (β) that could be evaluated independently of the question whether free willl and determinism were compatible [8].

---

7 For example, let W be any possible world. If 'N' is interpreted as 'it is true in W that', (β) is valid and N is not a classical necessity operator (as is easily seen from the fact that *p* could be false and N*p* true).

8 The best I know of is due to Thomas McKay. Suppose John has a choice about whether he plays

II

My allegiance to (β), therefore, is of quite independent of whatever opinions I may hold about Agglomeration and Closure, and an argument that led me to doubt one or both of these principles could very well leave my allegiance to (β) unshaken. (I cannot deny that an argument that led me to doubt Agglomeration or Closure could also be an argument that undermined my allegiance to (β). After all, any argument that did undermine my allegiance to (β) would almost certainly also be an argument that led me to doubt Agglomeration or Closure – since the invalidity of (β) entails that either Agglomeration or Closure is invalid. Does Slote say anything that one might aply «directly» to (β)? Only this. He displays various «plausible instances of alethic necessity» for which (β) fails [9]. For example: non-accidentality, irresistible impulse, and compulsion. These certainly seem to be in some sense types of «necessity», and Slote shows convincingly that (β) is not valid if 'N' is interpreted as expressing any of them. But what is the point of this procedure? I concede that it might serve to undermine an allegiance to (β) that was based on a belief that all necessity operators were «classical», or was based on an unexamined analogy between Choice Necessity and standard logical or metaphysical (or even physical) necessity. But I do not know of anyone whose allegiance to (β) *does* rest on so infirm a foundation. I know that mine doesn't, and I suspect that Ginet and Lamb and Wiggins would say the same.

I am, moreover, puzzled that Slote should bother discussing these relatively uninteresting selective necessity operators when it is easy to find selective necesstiy operators that seem to be far more relevant to the problem of free will and determinism. Consider this one, for example:

> *p* and no one is, or ever has been, such that if he were to choose to bring it about that it is (or was) false that *p*, then it would be (or would have been) false that *p*.

---

dice, and, in fact does play dice. But no one has a choice about how the dice fall in a fair game. The game *is* fair. John throws a six. The proposed counterexample is:

N John throws a six
N (John throws a six ⊃ Johns plays dice)
∴N John plays dice.

The first premise is true for the reason given. The second premise is true because the embedded conditional is a necessary truth. The conclusion is false for the reason given.

I reply: the first premise *is* false. John could have avoided throwing a six by avoiding playing dice. What John has no choice about is whether he throws a six given that he plays dice; that is, about whether (John plays dice ⊃ John throws a six).

9 I'm not sure what Slote means by 'alethic'. In a draft of this paper, I suggested «neither deontic nor doxastic», but Slote has denied (in correspondence) that this was what he meant. He did not, however, explain what he did mean.

The rule (β) fails for this operator[10]. Moreover, this operator is of special interest in discussions of the free will problem because adherents of the popular view that '*x* can do A' is equivalent to 'If *x* were to choose to do A, *x* would do A' would, presumably, say that it was equivalent to our Choice Necessity operator:

p and no one has, or ever had, any choice about whether *p*.

If this thesis – call it the Equivalence Thesis – is correct, it follows that the argument for incompatibilim that I laid out at the beginning of this paper is invalid. I am not greatly troubled by this. In fact, matters could hardly be otherwise. It is obvious that if 'can' statements are, as so many compatibilists allege, a certain sort of disguised conditional, then any argument for the incompatibility of free will and determinism is either invalid or else has false premises. It is obvious that if the Equivalence Thesis is correct, then 'can' statements are disguised conditionals of that type. It is obvious that our argument for incompatibilim has true premises. It is obvious that if our argument is invalid it is only because (β) is invalid. It is therefore not surprising that, if the Equivalence Thesis is correct, the (β) is invalid. So much the worse for the Equivalence Thesis, *I* say. It is obvious that (β) is valid (or so it seems to me); it is not obvious that the Equivalence Thesis holds; if two propositions are incompatible and one seems obviously true and the other does not seem obviously true, then, all other things being equal, one should accept the obvious member of the pair if one accepts either.

I am, therefore, unmoved by the fact that (β) fails if 'N' is interpreted as expressing the «were to choose» operator. But why then should I be moved by the fact that (β) fails if 'N' is interpreted as expressing the operators («it is no accident that *p*» and so on) that Slote calls our attention to? *I* do not think that the «were to choose» operator expresses Choice Necessity. But it is certainly true that many philosophers *have* thought that it does, or have held views that entail that it does. Therefore, there is an intimate and important connection, at least in the minds of many philosophers, between the 'were to choose' operator and the problem of free will and determinism. The operators that Slote displays, however, are much less intimately connected with the problem of free will and determinism, and the fact that (β) fails for these operators is consequently even less troubling to the incompatibilist that the fact that (β) fails for the «were to choose» operator.

Peter van Inwagen
Department of Philosophy
College of Arts and Sciences
541 Hall of Languages
Syracuse University
Syracuse
NEW YORK 13244-1170 (EE. UU.)

10 See *An Essay on Free Will*, p. 122 ff.