

## IV

THE INCOMPATIBILITY OF FREE WILL  
AND DETERMINISM\*

PETER VAN INWAGEN

IN this paper I shall define a thesis I shall call 'determinism', and argue that it is incompatible with the thesis that we are able to act otherwise than we do (i.e. is incompatible with 'free will'). Other theses, some of them very different from what I shall call 'determinism', have at least an equal right to this name, and, therefore, I do not claim to show that *every* thesis that could be called 'determinism' without historical impropriety is incompatible with free will. I shall, however, assume without argument that what I call 'determinism' is legitimately so called.

In Part I, I shall explain what I mean by 'determinism'. In Part II, I shall make some remarks about 'can'. In Part III, I shall argue that free will and determinism are incompatible. In Part IV, I shall examine some possible objections to the argument of Part III. I shall not attempt to establish the truth or falsity of determinism, or the existence or non-existence of free will.

## I

In defining 'determinism', I shall take for granted the notion of a proposition (that is, of a non-linguistic bearer of truth-value), together with

From *Philosophical Studies* 27 (1975), pp. 185-99.

Copyright © 1975 by D. Reidel Publishing Company, Dordrecht-Holland.  
Reprinted with permission.

\*The writing of this paper was supported by a stipend from the National Endowment for the Humanities for the summer of 1973. The paper was read at a colloquium at the University of Maryland at College Park. Earlier versions were read at the University of Rochester and Syracuse University. The audiences at these colloquia are thanked for useful comments and criticism. Special thanks are due to Rolf Eberle, Keith Lehrer, Raymond Martin, and Richard Taylor. I wish to thank Carl Ginet for his acute comments on an earlier draft, and the referee [of *Philosophical Studies*] for several helpful suggestions. Of course, none of these people is responsible for any mistakes that remain.

certain allied notions such as denial, conjunction, and entailment. Nothing in this paper will depend on the special features of any particular account of propositions. The reader may think of them as functions from possible worlds to truth-values or in any other way he likes, provided they have their usual features (e.g. they are either true or false; the conjunction of a true and a false proposition is a false proposition; they obey the law of contraposition with respect to entailment).

Our definition of 'determinism' will also involve the notion of 'the state of the entire physical world' (hereinafter, 'the state of the world') at an instant. I shall leave this notion largely unexplained, since the argument of this paper is very nearly independent of its content. Provided the following two conditions are met, the reader may flesh out 'the state of the world' in any way he likes:

(i) Our concept of 'state' must be such that, given that the world is in a certain state at a certain time, nothing follows *logically* about its states at other times. For example, we must not choose a concept of 'state' that would allow as part of a description of the momentary state of the world, the clause, '... and, at  $t$ , the world is such that Jones's left hand will be raised 10 seconds later than  $t$ .'

(ii) If there is some observable change in the way things are (e.g. if a white cloth becomes blue, a warm liquid cold, or if a man raises his hand), this change must entail some change in the state of the world. That is, our concept of 'state' must not be so theoretical, so divorced from what is observably true, that it be possible for the world to be in the *same* state at  $t_1$  and  $t_2$ , although (for example) Jones's hand is raised at  $t_1$  and not at  $t_2$ .

We may now define 'determinism'. We shall apply this term to the conjunction of these two theses:

- (a) For every instant of time, there is a proposition that expresses the state of the world at that instant.
- (b) If  $A$  and  $B$  are any propositions that express the state of the world at some instants, then the conjunction of  $A$  with the laws of physics entails  $B$ .

By a proposition that expresses the state of the world at time  $t$ , I mean a true proposition that asserts of some state that, at  $t$ , the world is in that state. The reason for our first restriction on the content of 'state' should now be evident: if it were not for this restriction, 'the state of the world' could be defined in such a way that determinism was trivially true. We could, without this restriction, build sufficient information about the past and future into each proposition that expresses the state of the world at an instant, that, for every

pair of such propositions, each *by itself* entails the other. And in that case, determinism would be a mere tautology, a thesis applicable to every conceivable state of affairs.

This amounts to saying that the 'laws of physics' clause on our definition does some work: whether determinism is true depends in the character of the laws of physics. For example, if all physical laws were vague propositions like 'In every nuclear reaction, momentum is *pretty nearly* conserved', or 'Force is *approximately* equal to mass times acceleration', then determinism would be false.

This raises the question, What is a law of physics? First, a terminological point. I do not mean the application of this term to be restricted to those laws that belong to physics in the narrowest sense of the word. I am using 'law of physics' in the way some philosophers use 'law of nature'. Thus, a law about chemical valences is a law of physics in my sense, even if chemistry is not ultimately 'reducible' to physics. I will not use the term 'law of nature', because, conceivably, *psychological* laws, including laws (if such there be) about the voluntary behaviour of rational agents, might be included under this term.<sup>1</sup> Rational agents are, after all, in some sense part of 'Nature'. Since I do not think that everything I shall say about laws of physics is true of such 'voluntaristic laws', I should not want to use, instead of 'laws of physics', some term like 'laws of nature' that might legitimately be applied to voluntaristic laws. Thus, for all that is said in this paper, it may be that some version of determinism based on voluntaristic laws is compatible with free will.<sup>2</sup> Let us, then, understand by 'law of physics' a law of nature that is not about the voluntary behaviour of rational agents.

But this does not tell us what 'laws of nature' are. There would probably be fairly general agreement that a proposition cannot be a law of nature unless it is true and contingent, and that no proposition is a law of nature if it entails the existence of some concrete individual, such as Caesar or the earth. But the proposition that there is no solid gold sphere 20 feet in diameter (probably) satisfies these conditions, though it is certainly not a law of nature.

It is also claimed sometimes that a law of nature must 'support its counterfactuals'. There is no doubt something to this. Consider, however, the proposition, 'Dogs die if exposed to virus V'. The claim that this proposition supports its counterfactuals is, I think, equivalent to the claim that 'Every

<sup>1</sup> For example, 'If a human being is not made to feel ashamed of lying before his twelfth birthday, then he will lie whenever he believes it to be to his advantage.'

<sup>2</sup> In 'The Compatibility of Free Will and Determinism', *Philosophical Review*, 1962, J. V. Canfield argues convincingly for a position that we might represent in this terminology as the thesis that a determinism based on voluntaristic laws could be compatible with free will.

dog is such that if it were exposed to virus V, it would die' is *true*. Let us suppose that this latter proposition *is* true, the quantification being understood as being over all dogs, past, present, and future. Its truth, it seems to me, is quite consistent with its being the case that dog-breeders *could* (but will not) institute a programme of selective breeding that *would* produce a sort of dog that is immune to virus V. But if dog-breeders *could* do this, then clearly 'Dogs die if exposed to virus V' is not a law of nature, since in that case the truth of the corresponding universally quantified counterfactual depends upon an accidental circumstance: if dog-breeders were to institute a certain programme of selective breeding they are quite capable of instituting, then 'Every dog is such that if it were exposed to virus V, it would die' would be false. Thus a proposition may 'support its counterfactuals' and yet not be a law of nature.

I do not think that any philosopher has succeeded in giving a (non-trivial) set of individually necessary and jointly sufficient conditions for a proposition's being a law of nature or of physics. I certainly do not know of any such set. Fortunately, for the purposes of this paper we need not know how to analyse the concept 'law of physics'. I shall, in Part III, argue that certain statements containing 'law of physics' are analytic. But this can be done in the absence of a satisfactory analysis of 'law of physics'. In fact, it would hardly be possible for one to *provide* an analysis of some concept if one had no pre-analytic convictions about what statements involving that concept are analytic.

For example, we do not have to have a satisfactory analysis of memory to know that 'No one can remember future events' is analytic. And if someone devised an analysis of memory according to which it was possible to remember future events, then, however attractive the analysis was in other respects, it would have to be rejected. The analyticity of 'No one can remember future events' is one of the *data* that anyone who investigates the concept of memory must take account of. Similarly, the claims I shall make on behalf of the concept of physical law seem to me to be basic and evident enough to be data that an analysis of this concept must take account of: any analysis on which these claims did not 'come out true' would be for that very reason defective.

## II

It seems to be generally agreed that the concept of free will should be understood in terms of the *power* or *ability* of agents to act otherwise than they in fact do. To deny that men have free will is to assert that what a man

does do and what he can do coincide. And almost all philosophers<sup>3</sup> agree that a necessary condition for holding an agent responsible for an act is believing that that agent *could have* refrained from performing that act.<sup>4</sup>

There is, however, considerably less agreement as to how 'can' (in the relevant sense) should be analysed. This is one of the most difficult questions in philosophy. It is certainly a question to which I do not know any non-trivial answer. But, as I said I should do in the case of 'law of physics', I shall make certain conceptual claims about 'can' (in the 'power' or 'ability' sense) in the absence of any analysis. Any suggested analysis of 'can' that does not support these claims will either be neutral with respect to them, in which case it will be incomplete, since it will not settle *all* conceptual questions about 'can', or it will be inconsistent with them, in which case the arguments I shall present in support of these claims will, in effect, be arguments that the analysis fails. In Part IV, I shall expand on this point as it applies to one particular analysis of 'can', the well-known 'conditional' analysis.

I shall say no more than this about the meaning of 'can'. I shall, however, introduce an idiom that will be useful in talking about ability and inability in complicated cases. Without this idiom, the statement of our argument would be rather unwieldy. We shall sometimes make claims about an agent's abilities by using sentences of the form:

S can render [could have rendered] . . . false.

where '. . .' may be replaced by names of propositions.<sup>5</sup> Our ordinary claims about ability can easily be translated into this idiom. For example, we translate:

He could have reached Chicago by midnight.

as

He could have rendered the proposition that he did not reach Chicago by midnight false.

and, of course, the translation from the special idiom to the ordinary idiom

<sup>3</sup> See, however, Harry Frankfurt, 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 1969.

<sup>4</sup> Actually, the matter is rather more complicated than this, since we may hold a man responsible for an act we believe he could not have refrained from, provided we are prepared to hold him responsible for his being unable to refrain.

<sup>5</sup> In all the cases we shall consider, '. . .' will be replaced by names of *true* propositions. For the sake of logical completeness, we may stipulate that any sentence formed by replacing '. . .' with the name of a *false* proposition is trivially true. Thus, 'Kant could have rendered the proposition that  $7 + 5 = 13$  false' is trivially true.

is easy enough in such simple cases. If we were interested only in everyday ascriptions of ability, the new idiom would be useless. Using it, however, we may make ascriptions of ability that it would be very difficult to make in the ordinary idiom. Consider, for example, the last true proposition asserted by Plato. (Let us assume that this description is, as logicians say, 'proper'.) One claim that we might make about Aristotle is that he could have rendered this proposition false. Now, presumably, we have no way of discovering *what* proposition the last true proposition asserted by Plato was. Still, the claim about Aristotle would seem to be either true or false. To discover its truth-value, we should have to discover under what conditions the last true proposition asserted by Plato (i.e. that proposition having as one of its accidental properties, the property of being the last true proposition asserted by Plato) would be false, and then discover whether it was within Aristotle's power to produce these conditions. For example, suppose that if Aristotle had lived in Athens from the time of Plato's death till the time of his own death, then the last true proposition asserted by Plato (whatever it was) would be false. Then, if Aristotle could have lived (i.e. if he had it within his power to live) in Athens throughout this period, he could have rendered the last true proposition asserted by Plato false. On the other hand, if the last true proposition asserted by Plato is the proposition that the planets do not move in perfect circles, then Aristotle could not have rendered the last true proposition asserted by Plato false, since it was not within his power to produce any set of conditions sufficient for the falsity of this proposition.<sup>6</sup>

It is obvious that the proposition expressed by 'Aristotle could have rendered the last true proposition asserted by Plato false', is a proposition that we should be hard put to express without using the idiom of rendering propositions false, or, at least, without using some very similar idiom. We shall find this new idiom very useful in discussing the relation between free will (a thesis about abilities) and determinism (a thesis about certain propositions).

### III

I shall now imagine a case in which a certain man, after due deliberation, refrained from performing a certain contemplated act. I shall then argue

<sup>6</sup> Richard Taylor has argued (most explicitly in 'Time, Truth and Ability' by 'Diodorus Cronus', *Analysis*, 1965) that every true proposition is such that, necessarily, no one is able to render it false. On my view, this thesis is mistaken, and Taylor's arguments for it can be shown to be unsound. I shall not, however, argue for this here. I shall argue in Part III that we are unable to render *certain sorts of* true proposition false, but my arguments will depend on special features of these sorts of proposition. I shall, for example, argue that no one can render false a law of physics; but I shall not argue that this is the case because laws of physics are *true*, but because of other features that they possess.

that, if determinism is true, then that man *could not have* performed that act. Because this argument will not depend on any features peculiar to our imagined case, the incompatibility of free will and determinism *in general* will be established, since, as will be evident, a parallel argument could easily be constructed for the case of any agent and any unperformed act.

Here is the case. Let us suppose there was once a judge who had only to raise his right hand at a certain time, *T*, to prevent the execution of a sentence of death upon a certain criminal, such a hand-raising being the sign, according to the conventions of the judge's country, of a granting of special clemency. Let us further suppose that the judge—call him '*J*'—refrained from raising his hand at that time, and that this inaction resulted in the criminal's being put to death. We may also suppose that the judge was unbound, uninjured, and free from paralysis; that he decided not to raise his hand at *T* only after a period of calm, rational, and relevant deliberation; that he had not been subjected to any 'pressure' to decide one way or the other about the criminal's death; that he was not under the influence of drugs, hypnosis, or anything of that sort; and finally, that there was no element in his deliberations that would have been of any special interest to a student of abnormal psychology.

Now the argument. In this argument, which I shall refer to as the 'main argument', I shall use '*T*<sub>0</sub>' to denote some instant of time earlier than *J*'s birth, '*P*<sub>0</sub>' to denote the proposition that expresses the state of the world at *T*<sub>0</sub>, '*P*' to denote the proposition that expresses the state of the world at *T*, and '*L*' to denote the conjunction into a single proposition of all laws of physics. (I shall regard *L* itself as a law of physics, on the reasonable assumption that if *A* and *B* are laws of physics, then the conjunction of *A* and *B* is a law of physics.) The argument consists of seven statements, the seventh of which follows from the first six:

- (1) If determinism is true, then the conjunction of *P*<sub>0</sub> and *L* entails *P*.
- (2) If *J* had raised his hand at *T*, then *P* would be false.
- (3) If (2) is true, then if *J* could have raised his hand at *T*, *J* could have rendered *P* false.<sup>7</sup>
- (4) If *J* could have rendered *P* false, and if the conjunction of *P*<sub>0</sub> and *L* entails *P*, then *J* could have rendered the conjunction of *P*<sub>0</sub> and *L* false.

<sup>7</sup> '*J* could have raised his hand at *T*' is ambiguous. It might mean either (roughly) '*J* possessed, at *T*, the ability to raise his hand', or '*J* possessed the ability to bring it about that his hand rose at *T*'. If *J* was unparalysed at *T* but paralysed at all earlier instants, then the latter of these would be false, though the former might be true. I mean '*J* could have raised his hand at *T*' in the latter sense.

- (5) If *J* could have rendered the conjunction of *P*<sub>0</sub> and *L* false, then *J* could have rendered *L* false.

- (6) *J* could not have rendered *L* false.

∴ (7) If determinism is true, *J* could not have raised his hand at *T*.

That (7) follows from (1) through (6) can easily be established by truthfunctional logic. Note that all conditionals in the argument except for (2) are truth-functional. For purposes of establishing the *validity* of this argument, (2) may be regarded as a simple sentence. Let us examine the premises individually.

- (1) This premiss follows from the definition of determinism.

(2) If *J* had raised his hand at *T*, then the world would have been in a different state at *T* from the state it was in fact in. (See our second condition on the content of 'the state of the world'.) And, therefore, if *J* had raised his hand at *T*, some contrary of *P* would express the state of the world at *T*. It should be emphasized that '*P*' does not *mean* 'the proposition that expresses the state of the world at *T*'. Rather, '*P*' *denotes* the proposition that expresses the state of the world at *T*. In Kripke's terminology, '*P*' is being used as a *rigid designator*, while 'the proposition that expresses the state of the world at *T*' is perforce non-rigid.<sup>8</sup>

(3) Since *J*'s hand being raised at *T* would have been sufficient for the falsity of *P*, there is, if *J* could have raised his hand, at least one condition sufficient for the falsity of *P* that *J* could have produced.

(4) This premiss may be defended as an instance of the following general principle:

If *S* can render *R* false, and if *Q* entails *R*, then *S* can render *Q* false.

This principle seems to be analytic. For if *Q* entails *R*, then the denial of *R* entails the denial of *Q*. Thus, any condition sufficient for the falsity of *R* is also sufficient for the falsity of *Q*. Therefore, if there is some condition that *S* can produce that is sufficient for the falsity of *R*, there is some condition (that same condition) that *S* can produce that is sufficient for the falsity of *Q*.

(5) This premiss may be defended as an instance of the following general principle, which I take to be analytic:

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.

<sup>8</sup> See Saul Kripke, 'Identity and Necessity', in *Identity and Individuation*, ed. Milton K. Munitz (New York, 1971).

Consider, for example, the propositions expressed by

The Spanish Armada was defeated in 1588.

and

Peter van Inwagen never visits Alaska.

The conjunction of these two propositions is quite possibly true. At any rate, let us assume it is true. Given that it is true, it seems quite clear that I can render it false if and only if I can visit Alaska. If, for some reason, it is not within my power ever to visit Alaska, then I *cannot* render it false. This is a quite trivial assertion, and the general principle (above) of which it is an instance is hardly less trivial. And it seems incontestable that premiss (5) is also an instance of this principle.

(6) I shall argue that if anyone *can* (i.e. has it within his power to) render some proposition false, then that proposition is not a law of physics. This I regard as a conceptual truth, one of the data that must be taken account of by anyone who wishes to give an analysis of 'can' or 'law'. It is this connection between these two concepts, I think, that is at the root of the incompatibility of free will and determinism.

In order to see this connection, let us suppose that both of the following are true:

(A) Nothing ever travels faster than light.

(B) Jones, a physicist, can construct a particle accelerator that would cause protons to travel at twice the speed of light.

It follows from (A) that Jones will never exercise the power that (B) ascribes to him. But whatever the reason for Jones's failure to act on his ability to render (A) false, it is clear that (A) and (B) are consistent, and that (B) entails that (A) is not a law of physics. For given that (B) is true, then Jones is able to conduct an experiment that would falsify (A); and surely it is a feature of any proposition that is a physical law that no one *can* conduct an experiment that would show it to be false.

Of course, most propositions that look initially as if they might be physical laws, but which are later decided to be non-laws, are rejected because of experiments that are actually performed. But this is not essential. In order to see this, let us elaborate the example we have been considering. Let us suppose that Jones's ability to render (A) false derives from the fact that he has discovered a mathematically rigorous proof that under certain conditions *C*, realizable in the laboratory, protons would travel faster than light. And let us suppose that this proof proceeds from premisses so obviously true that

all competent physicists accept his conclusion without reservation. But suppose that conditions *C* never obtain in nature, and that actually to produce them in the laboratory would require such an expenditure of resources that Jones and his colleagues decide not to carry out the experiment. And suppose that, as a result, conditions *C* are never realized and nothing ever travels faster than light. It is evident that if all this were true, we should have to say that (A), while *true*, is not a law of physics. (Though, of course, 'Nothing ever travels faster than light except under conditions *C*' might be a law.)

The laboratories and resources that figure in this example are not essential to its point. If Jones *could* render some proposition false by performing *any* act he does not in fact perform, even such a simple act as raising his hand at a certain time, this would be sufficient to show that that proposition is not a law of physics.

This completes my defence of the premisses of the main argument. In the final part of this paper, I shall examine objections to this argument suggested by the attempts of various philosophers to establish the compatibility of free will and determinism.

#### IV

The most useful thing a philosopher who thinks that the main argument does not prove its point could do would be to try to show that some premiss of the argument is false or incoherent, or that the argument begs some important question, or contains a term that is used equivocally, or something of that sort. In short, he should get down to cases. Some philosophers, however, might continue to hold that free will and determinism, in the sense of Part I, are compatible, but decline to try to point out a mistake in the argument. For (such a philosopher might argue) we have, in everyday life, *criteria* for determining whether an agent could have acted otherwise than he did, and these criteria determine the *meaning* of 'could have acted otherwise'; to know the meaning of this phrase is simply to know how to apply these criteria. And since these criteria make no mention of determinism, anyone who thinks that free will and determinism are incompatible is simply confused.<sup>9</sup>

As regards the argument of Part III (this philosopher might continue), this argument is very complex, and this complexity must simply serve to hide some error, since its conclusion is absurd. We must treat this argument

<sup>9</sup>Cf. Antony Flew, 'Divine Omniscience and Human Freedom', *New Essays in Philosophical Theology*, ed. Antony Flew and Alasdair MacIntyre (London: SCM Press, 1955), 149-51 in particular.

like the infamous 'proof' that zero equals one: It may be amusing and even instructive to find the hidden error (if one has nothing better to do), but it would be a waste of time to take seriously any suggestion that it is sound.

Now I suppose we do have 'criteria', in some sense of this over-used word, for the application of 'could have done otherwise', and I will grant that knowing the criteria for the application of a term can plausibly be identified with knowing its meaning. Whether the criteria for applying 'could have done otherwise' can (as at least one philosopher has supposed<sup>10</sup>) be taught by simple ostension is another question. However this may be, the 'criteria' argument is simply invalid. To see this, let us examine a simpler argument that makes the same mistake.

Consider the doctrine of 'predestinarianism'. Predestinarians hold (i) that if an act is foreseen it is not free, and (ii) that all acts are foreseen by God. (I do not claim that anyone has ever held this doctrine in precisely this form.) Now suppose we were to argue that predestinarianism must be compatible with free will, since our criteria for applying 'could have done otherwise' make no reference to predestinarianism. Obviously this argument would be invalid, since predestinarianism is incompatible with free will. And the only difference I can see between this argument and the 'criteria' argument for the compatibility of free will and determinism is that predestinarianism, unlike determinism, is *obviously* incompatible with free will. But, of course, theses may be incompatible with one another even if this incompatibility is not obvious. Even if determinism cannot, like predestinarianism, be seen to be incompatible with free will on the basis of a simple formal inference, there is, nonetheless, a conceptual connection between the two theses (as we showed in our defence of premiss (6)). The argument of Part III is intended to draw out the implications of this connection. There may well be a mistake in the argument, but I do not see why anyone should think that the very idea of such an argument is misconceived.

It has also been argued that free will *entails* determinism, and, being itself a consistent thesis, is *a fortiori* compatible with determinism. The argument, put briefly, is this. To say of some person on some particular occasion that he acted freely is obviously to say at least that *he* acted on that occasion. Suppose, however, that we see someone's arm rise and it later turns out that there was *no cause whatsoever* for his arm's rising. Surely we should have to say that *he* did not really raise his arm at all. Rather, his arm's rising was a mere chance happening, that, like a muscular twitch, had nothing to do with *him*, beyond the fact that it happened to involve a part of his body. A necessary condition for this person's really having raised his hand is that *he*

<sup>10</sup> Flew, loc cit.

caused his hand to rise. And surely '*he* caused' means '*his* character, desires, and beliefs caused'.<sup>11</sup>

I think that there is a great deal of confusion in this argument, but to expose this confusion would require a lengthy discussion of many fine points in the theory of agency. I shall only point out that if this argument is supposed to refute the conclusion of Part III, it is an *ignoratio elenchi*. For I did not conclude that free will is incompatible with the thesis that every event has a cause, but rather with determinism as defined in Part I. And the denial of this thesis does not entail that there are uncaused events.

Of course, one might try to construct a similar but relevant argument for the falsity of the conclusion of Part III. But, so far as I can see, the plausibility of such an argument would depend on the plausibility of supposing that if the present movements of one's body are not completely determined by physical law and the state of the world before one's birth, then these present movements are not one's own doing, but, rather, mere random happenings. And I do not see the least shred of plausibility in this supposition.

I shall finally consider the popular 'conditional analysis' argument for the compatibility of free will and determinism. According to the advocates of this argument—let us call them 'conditionalists'—what statements of the form:

(8) *S* could have done *X*

mean is:

(9) If *S* had chosen to do *X*, *S* would have done *X*.<sup>12</sup>

For example, 'Smith could have saved the drowning child' means, 'If Smith had chosen to save the drowning child, Smith would have saved the drowning child.' Thus, even if determinism is true (the conditionalists argue), it is possible that Smith did not save but *could have* saved the drowning child, since the conjunction of determinism with 'Smith did not save the child' does not entail the falsity of 'If Smith had chosen to save the child, Smith would have saved the child'.

Most of the controversy about this argument centres around the question whether (9) is a correct analysis of (8). I shall not enter into the debate about

<sup>11</sup> Cf. R. E. Hobart, 'Free Will as Involving Determination and Inconceivable Without It', *Mind*, 1934; A. J. Ayer, 'Freedom and Necessity', in his collected *Philosophical Essays* (New York, 1954) [Essay I in this collection.]; P. H. Nowell-Smith, 'Freewill and Moral Responsibility', *Mind*, 1948; J. J. C. Smart, 'Free Will, Praise, and Blame', *Mind*, 1961.

<sup>12</sup> Many other verbs besides 'choose' figure in various philosophers' conditional analyses of ability: e.g. 'wish', 'want', 'will', 'try', 'set oneself'. Much of the important contemporary work on this analysis, by G. E. Moore, P. H. Nowell-Smith, J. L. Austin, Keith Lehrer, Roderick Chisholm, and others, is collected in *The Nature of Human Action*, ed. Myles Brand (Glenview Ill., 1970). See also 'Fatalism and Determinism', by Wilfrid Sellars, in *Freedom and Determinism*, ed. Keith Lehrer (New York, 1966), 141–74.



whether this analysis is correct. I shall instead question the relevance of this debate to the argument of Part III. For it is not clear that the main argument would be unsound if the conditional analysis *were* correct. Clearly the argument is *valid* whether or not (8) and (9) mean the same. But suppose the premisses of the main argument were rewritten so that every clause they contain that is of form (8) is replaced by the corresponding clause of form (9)—should we then see that any of these premisses is false? Let us try this with premiss (6), which seems, *prima facie*, to be the crucial premiss of the argument. We have:

- (6a) It is not the case that if *J* had chosen to render *L* false, *J* would have rendered *L* false.

Now (6a) certainly seems true: If someone chooses to render false some proposition *R*, and if *R* is a law of physics, then surely he will fail. This little argument for (6a) *seems* obviously sound. But we cannot overlook the possibility that someone might discover a mistake in it and, perhaps, even construct a convincing argument that (6a) is false. Let us, therefore, assume for the sake of argument that (6a) is demonstrably false. What would this show? I submit that it would show that (6a) does not mean the same as (6), since (6) is, as I have argued, *true*.

The same dilemma confronts the conditionalist if he attempts to show, on the basis of the conditional analysis, that any of the other premisses of the argument is false. Consider the argument got by replacing every clause of form (8) in the main argument with the corresponding clause of form (9). If all the premisses of this new argument are true, the main argument is, according to the conditionalist's own theory, sound. If, on the other hand, any of the premisses of the new argument is false, then (*I* would maintain) this premiss is a counter-example to the conditional analysis. I should not be begging the question against the conditionalist in maintaining this, since I have given arguments for the truth of each of the premisses of the main argument, and nowhere in these arguments do I assume that the conditional analysis is wrong.

Of course, any or all of my arguments in defence of the premisses of the main argument may contain some mistake. But unless the conditionalist could point to some such mistake, he would not accomplish much by showing that some statement he *claimed* was equivalent to one of its premisses was false.<sup>13</sup>

<sup>13</sup> For an argument in some respects similar to what I have called the 'main argument', see Carl Ginet's admirable article, 'Might We Have No Choice?' in Lehrer, 87–104. Another argument similar to the main argument, which is (formally) much simpler than the main argument, but which is stated in language very different from that of traditional statements of the free-will problem, can be found in my 'A Formal Approach to the Problem of Free Will and Determinism', *Theoria*, 1974.

## V

## FREEDOM AND RESENTMENT

PETER STRAWSON

## I

SOME philosophers say they do not know what the thesis of determinism is. Others say, or imply, that they do know what it is. Of these, some—the pessimists perhaps—hold that if the thesis is true, then the concepts of moral obligation and responsibility really have no application, and the practices of punishing and blaming, of expressing moral condemnation and approval, are really unjustified. Others—the optimists perhaps—hold that these concepts and practices in no way lose their *raison d'être* if the thesis of determinism is true. Some hold even that the justification of these concepts and practices requires the truth of the thesis. There is another opinion which is less frequently voiced: the opinion, it might be said, of the genuine moral sceptic. This is that the notions of moral guilt, of blame, of moral responsibility are inherently confused and that we can see this to be so if we consider the consequences either of the truth of determinism or of its falsity. The holders of this opinion agree with the pessimists that these notions lack application if determinism is true, and add simply that they also lack it if determinism is false. If I am asked which of these parties I belong to, I must say it is the first of all, the party of those who do not know what the thesis of determinism is. But this does not stop me from having some sympathy with the others, and a wish to reconcile them. Should not ignorance, rationally, inhibit such sympathies? Well, of course, though darkling, one has some inkling—some notion of what sort of thing is being talked about. This lecture is intended as a move towards reconciliation; so is likely to seem wrongheaded to everyone.

But can there be any possibility of reconciliation between such clearly

From *Proceedings of the British Academy*, vol. xlviii (1962), pp. 1–25. Reprinted by permission of the British Academy.