

The Consequence Argument

Peter van Inwagen

In a book I once wrote about free will, I contended that the best and most important argument for the incompatibility of free will and determinism was “the Consequence Argument.” I gave the following brief sketch of the Consequence Argument as a prelude to several more careful and detailed statements of the 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.

The reading that follows this one, Reading 41, “The Mystery of Metaphysical Freedom,” contains a statement of the Consequence Argument. The argument is contained in the paragraph (p. xxx) that starts, “As Carl Ginet has said” But, as you will see if you compare the “brief sketch” with that paragraph, “The Mystery of Metaphysical Freedom” presents the Consequence Argument in a disguise that is not easy to penetrate. Some teachers of philosophy who have used the first edition of Metaphysics: The Big Questions as a textbook have asked for a more straightforward statement of the Consequence Argument (since much of the recent discussion of the question of the compatibility of free will and determinism in the philosophical literature has taken the form of criticisms of the Consequence Argument that are rather hard to apply to the argument in the form in which it is presented in Reading 41). This essay is an attempt to meet this request.

Some truths, some true propositions, have this important feature (well, it’s important to me): their truth, the fact that they are true, is something that is or once was up to me. For example, the proposition that the title this essay is “The Consequence Argument” is true, and it was once up to me whether it would be true. Or, at any rate, almost everyone would suppose that it was once up to me whether it would be true. If we believe that the truth of this proposition is something that is up to me, we believe this because we believe that (in this matter, at any rate) I have free will: that it is, or once was, up to me what title the essay would have, that I was able to give to give it any of various other appropriate titles. If I have free will, in short, then some things are up to me. (Even the outcome of a process that is not under my control can be up to me, provided it was up to me whether the process occurred at all. I cannot control the way a fair die falls, but the truth of the true proposition “The die fell ‘four’” was up to me if it was up to me whether the die was thrown at all.) Some things are up to me—but not, of course, all things, for free will does not imply omnipotence. I was, for example, born in 1942, and whether I was born in that year is not, and never was, up to me. If, however, my parents had free will, the truth of this proposition is something that was up to them: if they had chosen not to marry or never to have children or to wait to have children till my father was released from military service (which happened in 1944), then the proposition that I was born in 1942 would have been false.

Some truths, however, are not up to anyone (to any human being). For example: that human beings exist, that the earth has a large moon, that the presence of mass changes the local curvature of spacetime, that there is no largest prime number. Let us call a true proposition whose truth is up to no one (to no human being, past, present, or future) an untouchable proposition. (We shall sometimes find it convenient to say that it is an untouchable fact that or an untouchable truth that, e.g., the earth has a large moon.)

An untouchable proposition has the following feature: it is true and nothing that anyone is or ever has been able to do (no possible combination or series of actions that any human being has been able to perform individually or that any human beings have been able to perform collectively) would have had the consequence that it was false. It might even be supposed that this would be another way of saying that an untouchable proposition is a true proposition whose truth is up to no one. And perhaps this supposition would be consistent with the meaning of “true proposition whose truth is up to no one.” However that may be, I mean something stronger than this by “untouchable proposition” or “true proposition whose truth is up to no one.” I mean this:

An untouchable proposition is a true proposition that is such that nothing that anyone is or ever has been able to do might have had the consequence that it was false.

To see the difference between the (weaker) “would have had” and the (stronger) “might have had” understanding of ‘untouchable’, consider this case. Suppose that no human being is ever invisible. And suppose I once had a single vial of magical potion (which I was able to drink but in fact did not drink—in fact I chose to pour it down the drain) with this property: if I had drunk it I might have become invisible—and I might not have. Let us say, in fact, that if I had drunk the potion in the vial, there was a fifty-fifty chance that my ingesting it would have resulted in my becoming invisible. And let us say that drinking the potion is the only thing I have (or anyone has) ever been able to do that might have had the consequence that I (or anyone else) became invisible. If all this is true, then the true proposition “No human being is ever invisible” is not, as I have defined the term, an untouchable proposition. Although nothing I have ever been able to do would have had the consequence that this proposition was false, something I was once able to do might have had the consequence that this proposition was false; and that is sufficient for this proposition’s not being untouchable.

Untouchability seems to have a certain logic to it. One part of this logic is this: any necessary truth (any proposition that has to be true, that would be true no matter what, like “2+2=4” or “An airplane can’t be made entirely out of gaseous hydrogen”) is untouchable. Let us call the following rule of logical inference:

It is a necessary truth that p

hence, it is an untouchable truth that p

the Necessity Rule. Another part of the logic of untouchability is captured in a rule about conditional (if-then) statements:

It is an untouchable truth that p

It is an untouchable truth that (if p , then q)

hence, it is an untouchable truth that q .

Let us call this the Conditional Rule. Examination of a few examples will, I think, convince the reader that it is at least very plausible to suppose that the Conditional Rule is valid. Consider, for instance, the result of replacing ' p ' in the Conditional Rule with 'The sun explodes in the year 2027' and ' q ' with 'All life on the earth ends in the year 2027':

It is an untouchable truth that the sun explodes in the year 2027

It is an untouchable truth that (if the sun explodes in the year 2027, then all life on the earth ends in the year 2027)

hence, it is an untouchable truth that all life on the earth ends in the year 2027'.

If in fact the sun does explode in 2027, then the first premise is no doubt true. (If the sun is going to explode in 2027, then nothing human beings can do would—or might—prevent that catastrophic event.) And the second premise seems very plausible, too: If the sun explodes, then nothing human beings can do would—or might—prevent the immediate extinction of all terrestrial life. Whether the two premises of this argument are true or not, however, it seems evident beyond all possibility of dispute that the conclusion of the argument follows from them: if we can't prevent the explosion of the sun, and if we can't do anything about the fact that such an explosion would have the end of life on the earth as a consequence, then we can't prevent the end of all life on the earth. The reader is invited to try to construct a counterexample to the Consequence Rule. That is, to imagine a possible case in which, for some argument that comes from the rule by substitution of any sentences for ' p ' and ' q ', the premises of the argument are both true and its conclusion false. It is at least very hard to find a counterexample to the Consequence Rule. But the rule is invalid only if there is some possible counterexample to it.

Having set out these two rules governing the notion of untouchability, let us turn to the idea of determinism. Let P_0 be a proposition that gives a complete and correct description of the state of the whole universe at some time in the remote past (a million years ago, say). And let L be the conjunction into a single proposition of all the laws of nature (or the laws of physics). (By the laws of nature we do not mean what physicists or other scientists now happen to think the laws of nature are. We mean the real laws of nature, the laws as God sees them, or as a complete and perfect science at the end of all enquiry would see them.) Determinism implies that the following conditional (if-then) proposition is a necessary truth:

(1) If P_0 and L are both true, van Inwagen writes an essay called "The Consequence Argument."

For suppose that determinism is true. Determinism says that the past (the past at any given instant, a complete specification of the universe at any given instant in the past) and the laws of nature together determine everything, that they leave no open possibilities whatever. And since van Inwagen did write an essay called "The Consequence Argument," the fact that he wrote this essay is one aspect of the "everything" that is determined by the past and the laws.

The laws of elementary logic tell us that (1) is logically equivalent to

(2) If P_0 is true, then, if L is true, van Inwagen writes an essay called "The Consequence Argument."

If (1) is a necessary truth (as we have seen that it is), then (2), being logically equivalent to (1), must also be a necessary truth. And if (2) is a necessary truth, it is, by the Necessity Rule, an untouchable truth.

Now consider P_0 . P_0 is obviously an untouchable truth. P_0 is an untouchable truth for the same reason that "Dinosaurs once walked the earth" is an untouchable truth: both are truths about the past, and, indeed, truths about the pre-human past.

Let us now examine the following argument:

It is an untouchable truth that P_0 is true

It is an untouchable truth that (if P_0 is true, then, if L is true, van Inwagen writes an essay called "The Consequence Argument")

hence, it is an untouchable truth that if L is true, van Inwagen writes an essay called "The Consequence Argument".

Examination shows that this argument is valid if the Conditional Rule is valid. (It comes from the Conditional Rule by substitution of ' P_0 is true' for ' p ' and 'if L is true, van Inwagen writes an essay

called "The Consequence Argument" for 'g'. And the two premises of the argument are, as we have seen, true. (Note that what follows 'it is an untouchable truth that' in the second premise is just statement (2).) The conclusion of the argument is therefore true (given that the Conditional Rule is valid).

Now examine a second argument:

It is an untouchable truth that L is true

It is an untouchable truth that (if L is true, van Inwagen writes an essay called "The Consequence Argument")

hence, it is an untouchable truth that van Inwagen writes an essay called "The Consequence Argument".

This argument, too, comes from the Conditional Rule by substitution: of 'L is true' for 'p' and of 'Van Inwagen writes an essay called "The Consequence Argument"' for 'q'. The first premise of this second argument is true, for L is the conjunction into a single proposition of all the laws of nature. And the laws of nature are untouchable truths. As far as human beings (at any rate) are concerned, the laws of nature are just there, one of the givens of our existence. nothing anyone is able to do is such that it would, or even might, result in the falsity of a proposition that is a law of nature. (If there is an experiment that physicists are able to perform—even if they never do—that might result in a violation of the principle of the conservation of angular momentum, it follows that the principle of the conservation of angular momentum is not a law of nature.) And the second premise is the conclusion of the first argument. The conclusion of the argument is therefore true (if, again, the Conditional Rule is valid).

What does the conclusion say? Well it certainly implies that I have, and never had, any free will in the matter of whether I should write an essay called "The Consequence Argument." It implies that I was never able not to write an essay of that title: for if I did have that ability, then there is something I was able to do such that, if I had done it, the proposition "Van Inwagen writes an essay called "The Consequence Argument" would have been false. If I did have that ability, then that proposition would not have been an untouchable proposition.

This is, of course, a rather limited conclusion. It says nothing about anyone's free will but mine, and it leaves plenty of room for free will for me in matters unrelated to the above essay and its title. But it is obvious that the argument we have gone through is easily generalized to show that no one has any free will in any matter whatever. For, if there is no error in the above argument, then there will be no error in any argument that is obtained from it by replacing each occurrence of 'Van Inwagen writes an essay called "The Consequence Argument"' with any sentence that expresses a truth. Appropriate substitution, for example, will yield arguments whose conclusions are:

It is an untouchable truth that James Earl Ray assassinated Martin Luther King, Jr.

It is an untouchable truth that the World Trade Center was destroyed by terrorists on September 11th, 2001.

It is an untouchable truth that six million Jews were murdered in the Holocaust.

And this implies that if determinism is true (if the past and the laws of nature determine a unique future), every true proposition is an untouchable truth. And if every true proposition is an untouchable truth, then free will simply does not exist: no one is ever able to anything other than just exactly those things that he or she does.

This conclusion, which is called incompatibilism, is a very strong thesis indeed. Is there any way to avoid it? Anyone who wishes to avoid it—anyone who wishes to be a compatibilist, anyone who wishes to believe that free will and determinism can co-exist—must deny at least one of the following propositions, each of which the argument depends on in one way or another:

The Necessity Rule is valid

The Conditional Rule is valid

P₀ is an untouchable truth

L is an untouchable truth.

Or, at any rate, this is the case if the meanings of all four propositions (or the meanings of the sentences that express them) are perfectly clear. The philosopher David Lewis has contended (this is a translation of what he has contended into the terms used in this essay) that our technical term 'untouchable' is ambiguous, and that if the word is understood in one of its possible senses, the Conditional Rule is invalid, and, if it is understood in the other of its possible senses, L is not an untouchable truth. I will count this as a special case of denying that all four propositions are true.

Most critics of the argument agree that its weak point is the validity of the Conditional Rule. Most of its defenders would concede that if the argument has a weak point, that weak point is the validity of the Conditional Rule. It would seem, therefore, that the lesson of the Consequence Argument is that the question of the compatibility of free will and determinism, in the last analysis, comes down to the question whether the Conditional Rule is valid.

References

Peter van Inwagen, An Essay on Free Will (Oxford: at the Clarendon Press, 1983), p. 16 and p. 56.

David Lewis, "Are We Free to Break the Laws," Theoria 47 (1981), pp. 113-21. Reprinted in David Lewis, Philosophical Papers, Vol. II (New York and Oxford: Oxford University Press, 1986), pp. 291-98. See also Peter van Inwagen, "Freedom to Break the Laws," Midwest Studies in Philosophy Vol. XXVIII (2004), pp. 334-50.