The Philosophical Review, Vol. 109, No. 4 (October 2000)

Van Inwagen's Consequence Argument

Michael Huemer

Peter van Inwagen has presented a compelling argument for the incompatibility of free will and determinism, which he calls "the Consequence Argument."[1] This argument depends on a controversial inference rule, "rule beta," which says

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

In van Inwagen's view, as well as my own, the Consequence Argument is the strongest argument for incompatibilism, and, as he formulates the argument, rule beta is its weakest link.[2] For this reason, it is of the utmost importance, in the debate over compatibilism, to determine whether rule beta is valid.

In the following, I show, first, that given van Inwagen's definition of 'Np', rule beta is invalid. Its invalidity can be demonstrated by use of a counterexample that does not presuppose either determinism or compatibilism. Second, however, I identify an alternative definition of 'Np' that would suit van Inwagen's purposes, and show that under this alternative definition, rule beta is valid. The rest of van Inwagen's argument remains highly plausible when this alternative definition is used.

1. The Consequence Argument

1.1 Van Inwagen's Formulation

The Consequence Argument makes use of a sentential operator, 'N', which is defined as follows:

I would like to thank Michael Tooley, Graham Oddie, and the reviewers at the *Philosophical Review* for their helpful comments on the manuscript. I am also indebted to Adam Vinueza for a conversation that clarified and sharpened for me the argument of the last section of this paper.

[1]Peter van Inwagen, *An Essay on Free Will* (Oxford: Oxford University Press, Clarendon Press, 1983), chapter 3. Three different versions of the argument appear there, of which I shall focus on the third. Van Inwagen has indicated that the latter is his favorite version ("When Is the Will Free?" *Philosophical Perspectives* 3 (1989): 399–422, at 404) and that, in his opinion, the three arguments will stand or fall together (*Essay on Free Will,* 56–57).

[2]Van Inwagen, *Essay on Free Will,* 96; "When Is the Will Free?" 405.

Np = No one has any choice about the fact that p.[3]

Van Inwagen puts forward two inference rules involving 'N':

Rule α: From $\Box p$, deduce Np (where '$\Box$' represents logical necessity).

Rule β: From Np and $N(p \supset q)$, deduce Nq.

He then argues as follows. Let P_0 be a true proposition describing the complete state that the universe was in at some instant in the remote past (in, say, 4 billion B.C.). Let L be the conjunction into a single proposition of all the laws of nature. And let P be a true proposition describing any occurrence after the time at which the state described by P_0 held. If determinism is true, then P follows from the conjunction of P_0 and L. Assume, for conditional proof, that determinism is true. Then we can argue:

1. $\Box((P^0 \,\&\, L) \supset P)$	assumption
2. $\Box(P_0 \supset (L \supset P))$	1; modal & sentential logic
3. $N(P_0 \supset (L \supset P))$	2; rule α
4. NP_0	premise
5. $N(L \supset P)$	3, 4; rule β
6. NL	premise
7. NP	5, 6; rule β[4]

This appears to show that if determinism is true, then no one has or ever had any choice about anything that happened in human history, nor about anything that ever will happen.

Call this the First Version of the Consequence Argument.

1.2 An Alternative Formulation

The argument might also be formulated even more simply in the following way. Instead of rules α and β above, we use:

---

[3]Van Inwagen, *Essay on Free Will,* 93. Note that Np, so defined, entails p. I have abbreviated van Inwagen's definition slightly, but not so as to alter its sense. 'Has' in this definition may be read as timeless.

[4]Van Inwagen, *Essay on Free Will,* 93–95, and "When Is the Will Free?" 404–5.

Rule α*: From Np and $p \rightarrow q$, deduce Nq (where '$\rightarrow$' represents entailment).

Rule β*: From Np and Nq, deduce N(p & q).[5]

Assume again that determinism is true. Then we can argue:

1.	$(P_0 \ \& \ L) \rightarrow P$	assumption
2.	NP_0	premise
3.	NL	premise
4.	$N(P_0 \ \& \ L)$	2, 3; rule β*
5.	NP	1, 4; rule α*

Call this the Second Version of the Consequence Argument.

It can be shown that (modulo the assumption that there is at least one thing we have no choice about) the First Version of the Consequence Argument is valid if and only if the Second Version is valid. That is to say, it can be shown that rules α and β are both valid, if and only if rules α* and β* are both valid. To see this, assume that X is any fact about which we have no choice, and:

I. Assume α and β are both valid. Then

A. α* is valid. For

1. Suppose we have Np and $p \rightarrow q$. Then:

2.	$\Box(p \supset q)$	1
3.	$N(p \supset q)$	2; rule α
4.	Nq	1, 3; rule β

[5]Michael Slote ("Selective Necessity and the Free-Will Problem," *Journal of Philosophy* 79 (1982): 5–24, especially 10) argues that the Consequence Argument depends on these two rules for its plausibility. André Gallois ("Van Inwagen on Free Will and Determinism," *Philosophical Studies* 32 (1977): 99–105, 101) has also argued that the Consequence Argument depends on rule β*. These same rules figure in James Lamb's version of the Consequence Argument ("On a Proof of Incompatibilism," *Philosophical Review* 86 (1977): 20–35): rule α* corresponds to his principle of "can-entailment" (23), while rule β* corresponds to his "union inefficacy" principle (24). See also note 11 below.

B. β* is also valid. For

1. Suppose we have Np and Nq.
2. $\Box(p \supset (q \supset (p \,\&\, q)))$ logical truth
3. $N(p \supset (q \supset (p \,\&\, q)))$ 2; rule α
4. $N(q \supset (p \,\&\, q))$ 1, 3; rule β
5. $N(p \,\&\, q)$ 1, 4; rule β[6]

II. Assume α* and β* are both valid. Then

A. α is valid. For

1. Suppose we have $\Box p$.
2. NX by assumption
3. $X \rightarrow p$ 1
4. Np 2, 3; rule α*

B. β is valid. For

1. Suppose we have Np and $N(p \supset q)$.
2. $N(p \,\&\, (p \supset q))$ 1; rule β*
3. $(p \,\&\, (p \supset q)) \rightarrow q$ logical truth
4. Nq 2, 3; rule α*

Thus, if we can successfully criticize any of the rules—α, β, α*, or β*—we will thereby show that both versions of the consequence argument fail.[7]

Now, in order to determine whether these inference rules are valid, we need to know more about what it means to "have a choice about" a proposition. In the next three sections, I consider three different ways of understanding this notion. The first two derive from van Inwagen; the third is my own.

---

[6]The proof of β* also appears in Thomas McKay and David Johnson, "A Reconsideration of an Argument against Compatibilism," *Philosophical Topics* 24 (1996): 113–22, at 115.

[7]This is true provided, of course, that we do not rely on a counterexample to α in which X doesn't exist. That is, if we were to criticize α by appealing to a possible world in which every proposition is one that someone has a choice about, such a criticism would not tell against the Second Version of the Consequence Argument. But of course, no such criticism is in the offing here.

2. The Rendering-False Interpretation

According to van Inwagen, to say S has a choice about the fact that p is the same as to say that although p is the case, S *can render* p *false.* This, in turn, is equivalent to saying that although p, there is some action S can perform, his performance of which is *sufficient* for $\sim p$.[8]

What kind of "sufficiency" is involved here? Van Inwagen specifies *logical* sufficiency. At first glance, this seems to be a mistake, as it will give a far too restrictive account of what it is to have a choice about something. For instance, I might have a choice about the fact that Sara's phone won't ring at 11:00 tonight by virtue of the facts that I can call her up at 11:00 and that my doing so would cause her phone to ring at 11:00 tonight—even though my calling at this time would not be *logically* sufficient for the phone's ringing at this time. Van Inwagen could respond to this problem by allowing the (causal) consequences of an action to be built into a description of the action. Thus, it might be said that one of the actions I can perform is that of *making Sara's phone ring at 11:00,* my doing so being logically sufficient for Sara's phone to ring at 11:00, and that it is by virtue of this that I have a choice about whether Sara's phone rings at 11:00.

A second possibility is to use *counterfactual* sufficiency: that is, S can render p false (where p is some true proposition) iff S can perform some action, such that if he were to do so, it would not be the case that p.[9]

A third possibility is this: S can render p false (where p is some true proposition) iff S can perform some action, such that if he were to do so, S would *thereby bring it about* that $\sim p$.[10]

For his argument to succeed, van Inwagen need only find *some*

---

[8]Van Inwagen, "Reply to Narveson," *Philosophical Studies* 32 (1977): 89–98, at 93, and *Essay on Free Will,* 66–67. For the sake of simplicity, I omit van Inwagen's later qualification that requires holding the past fixed (*Essay on Free Will,* 68). This qualification does not affect the counterexamples in the following two subsections.

[9]This corresponds to David Lewis's "weak sense" of 'can render false' ("Are We Free to Break the Laws?" *Theoria* 47 (1981): 113–21, at 120).

[10]This proposal derives from Thomas Talbott, "On Free Agency and the Concept of Power," *Pacific Philosophical Quarterly* 69 (1988): 241–54, at 251. It also appears to be equivalent to Lewis's "strong sense" of 'can render false' ("Are We Free to Break the Laws?" 120). Lewis suggests (116–17) that, if A is some action S could have but did not perform, then if S *had* done

interpretation of 'can render false' under which rule beta is valid, the premises of the Consequence Argument remain plausible, and the universal closure of Np implies the absence of free will. Unfortunately, however, none of the above three interpretations of 'can render false' succeeds. A single example can be used to show the invalidity of rules β and β* on any of these three versions of the "rendering false" interpretation.

*2.1 A Counterexample to β**

Consider the following case.

> S has free will and lives in an indeterministic world. S has access to a certain device that shoots R-particles into an R-particle basket. Given the current setup of the device, the basket, and their immediate environment, it is nomologically necessary that, if and only if the device is activated, an R-particle will be shot into the basket. However, the laws of nature do *not* determine, in the given setup, whether the R-particle would land in the left half of the basket or the right half. Which half the particle landed in, the laws say, would be a matter of chance—there being, if you like, a 50% probability of each; nevertheless, it is determined that the R-particle would definitely land in the basket. Neither S nor anyone else can alter the setup of the device in any way that would affect these facts. Whether the device is activated or not is determined by S's free choice. S decides not to activate the device, and no R-particle is emitted.

Now consider the propositions,

> A = No R-particle lands in the left half of the basket.
> B = No R-particle lands in the right half of the basket.

A and B are both true in the above scenario. Furthermore, S cannot render A false, under any of our three readings of 'can render false'. First, S cannot do anything that would be logically sufficient

---

A, an (actual) law of nature would have been violated; however, S would not have *brought it about* that the law was violated, since the law-violation would have happened before S did A. Thus, according to Lewis, S can render a law of nature false in the weak sense but not in the strong sense. In other words, according to Lewis, S would have a choice about the fact that L on our second rendering-false interpretation but not on the third.

for ~A (that is, logically sufficient for an R-particle to land in the left half of the basket). The only relevant thing S can do is to activate the device, which obviously does not logically guarantee that an R-particle lands in the left half of the basket. If we allow the would-be causal consequences of this action to be built into descriptions of the action, then we can also say: S can shoot an R-particle into the basket, and S can make it 50% likely that an R-particle lands in the left half of the basket. But neither of these things is logically sufficient for an R-particle's landing in the left half of the basket either. Second, S cannot do anything that would be counterfactually sufficient for ~A—it is not the case that if S were to activate the device, then an R-particle *would* land in the left half of the basket; rather, if S were to activate the device, an R-particle *might or might not* land in the left half of the basket. That is, in some of the nearby possible worlds in which S activates the device the R-particle enters the left half of the basket, and in some of the nearby worlds in which S activates the device the R-particle enters the right half instead. Third, *a fortiori* it is not the case that if S were to activate the device, then S would bring it about that an R-particle landed in the left half of the basket. Thus, S cannot render A false, for S cannot in any sense *guarantee* that an R-particle enters the left half of the basket.

By parity of reasoning, S cannot render B false either.

However, S *can* render the conjunction of A and B false. The conjunction of A and B is simply the proposition that no R-particle enters the basket, and S can render this false in all three senses of 'can render false'. First, S can activate the device and thereby shoot an R-particle into the basket, and S's shooting an R-particle into the basket is logically sufficient for an R-particle's landing in the basket (hence, logically sufficient for ~(A & B)). Second, S can activate the device, and if S were to do so, it would not be the case that no R-particle entered the basket. And third, if S were to activate the device, S would thereby bring it about that an R-particle entered the basket (hence, bring it about that ~(A & B)).

Thus, we have

NA,
NB, but
~N(A & B),

which is a counter-example to rule β*.[11]

The essential method of criticizing rule β*, then, is this: come up with an action that is sufficient for $(p \vee q)$, but which is not sufficient for p and is not sufficient for q. Specify the example so that this is the only relevant action S can perform. Then argue that S cannot render $\sim p$ false, cannot render $\sim q$ false, but *can* render $(\sim p \ \& \ \sim q)$ false. This will refute rule β*, provided that 'Np' is defined in terms of rendering p false. The reason why no interpretation of 'sufficient' will validate rule β* is simply that there is no kind of sufficiency such that, from the fact that r is sufficient for $(p \vee q)$, it follows that r must be either sufficient for p or sufficient for q.[12] Logical sufficiency clearly doesn't satisfy this condition, and the possibility of indeterminism prevents either counterfactual sufficiency or causal sufficiency from satisfying it either. And given the invalidity of β*, β must also be invalid.[13]

2.2 A Counterexample to β

Given the above example, it is easy now to construct the relevant counterexample to rule β. Consider the same scenario as above; however, this time, consider the following two propositions:

---

[11]This counterexample strongly resembles those brought forth by Gallois, "Van Inwagen on Free Will and Determinism," 101, and McKay and Johnson, "A Reconsideration," 115. Van Inwagen ("Reply to Gallois," *Philosophical Studies* 32 (1977): 107–11), however, seems to miss the significance of Gallois's examples, in part because they were brought forth in response to a different version of the Consequence Argument from those discussed here. (Gallois is responding to van Inwagen, "The Incompatibility of Free Will and Determinism," *Philosophical Studies* 27 (1975): 185–99, which does not make any explicit mention of rules α, β, α*, or β*.)

[12]This point is made by Terence Horgan ("Compatibilism and the Consequence Argument," *Philosophical Studies* 47 (1985): 339–56, at 341–42). Horgan, like Gallois and Slote, underestimates the strength of his criticism—all three authors think that the invalidity of rule β* *undermines* rule β, removing the principal or only reason for accepting β. In fact, as McKay and Johnson realize (115–16), the invalidity of rule β* *proves* the invalidity of either α or β.

[13]McKay and Johnson have deployed a similar example, involving coin-flipping, against rule β*, although theirs does not involve genuinely random events. They claim (118) that it is an advantage of their example that it does not stipulate indeterminism. But note that genuine randomness is required to refute β under the counter-factual interpretation of 'Np', since, given determinism, the counterfactual, 'If I flipped the coin at t, it would have come up heads' is (arguably, at least) determinate. We need an example in which, if S had made a certain choice, either p or q would definitely have held, but it is indeterminate which one would have held.

A = No R-particle lands in the left half of the basket.
C = No R-particle lands in the basket.

Once again, S has no choice about the fact that A, since he cannot render A false. In addition, S has no choice about the fact that (A ⊃ C). In order for S to be able to render (A ⊃ C) false, S must be able to do something, his doing which is sufficient for ~(A ⊃ C). Now, ~(A ⊃ C) is logically equivalent to (A & ~C), which is logically equivalent to ~B (if no particle lands in the left half of the basket, but a particle does land in the basket, it must land in the right half). S cannot do anything sufficient for (anything that would *guarantee*) the particle's landing in the right half of the basket. Therefore, S cannot do anything, his doing which is sufficient for ~(A ⊃ C). Therefore, S cannot render (A ⊃ C) false.

However, S does have a choice about the fact that C, for, again, S can shoot an R-particle into the basket, and his doing so would be sufficient for ~C. Thus, we have

NA,
N(A ⊃ C), but
~NC,

which is a counterexample to β.[14]

The above criticism of rule β has an important virtue that is worth noticing: it does not depend on assuming that compatibilism is or might be true. Several other authors have made criticisms that do depend on such an assumption. In essence, they have argued that van Inwagen's argument begs the question against compatibilism, because if compatibilism is true, then some step of van Inwagen's argument is clearly mistaken (though they do not all agree on which step).[15] Van Inwagen has, in my opinion, adequately rebutted this kind of charge. It is simply a consequence of the fact that van Inwagen's argument is an argument *against* compatibilism that, if compatibilism is true, then some step of his argument must

[14]See also David Widerker, "On an Argument for Incompatibilism," *Analysis* 47 (1987): 37–41, for two similar counterexamples to β.

[15]I have in mind such authors as Richard Foley ("Compatibilism and Control over the Past," *Analysis* 39 (1979): 70–74), Christopher Hill ("Van Inwagen on the Consequence Argument," *Analysis* 52 (1992): 49–55), David Lewis ("Are We Free to Break the Laws?" 119 n.), and Thomas Flint ("Compatibilism and the Argument from Unavoidability," *Journal of Philosophy* 84 (1987): 423–40).

be mistaken; this does not mean his argument begs the question, unless all deductive arguments beg the question. Nor does the fact that his argument has a clear weakest link (if this is a fact) make it question-begging.[16] Accordingly, my criticism of rule β, above, does not say that rule β is invalid provided determinism or compatibilism is true. My example, rather, appeals to a world in which free will exists and indeterminism holds—both of which van Inwagen believes to be true in the actual world. Thus, I have shown that even if libertarianism is true, rule β is still invalid.

3. The Possible-Worlds Interpretation

Van Inwagen suggests another way of interpreting 'Np':

Np = p is true in all the possible worlds that S has access to.

Intuitively, whenever a person can bring it about that p, he may be said to "have access to" a world in which p is true. S's having free will may be thought of as S's having access to some possible worlds other than the actual world.[17] Van Inwagen points out that if 'Np' can be understood in this way—no matter what the set of worlds I have access to is—then rule β will be valid. (If p holds in some set of worlds, and $(p \supset q)$ also holds in that set of worlds, then q holds in that set of worlds.)[18] Thus, to validate β, all van Inwagen need do is find a set of possible worlds that could reasonably be thought of as the set of worlds S has access to.

The notion of having access to a world is a technical notion, so it is up to van Inwagen to define it as he chooses. One pitfall must be carefully avoided, however: it is possible that once the set of worlds S has access to is delineated, it will no longer be plausible that "having access" to possible worlds has anything to do with free will. Here is a trivial example of the potential problem: suppose we define the worlds S has access to to be just the nomologically possible worlds. In that case, it will seem that my having access to a possible but non-actual world has little to do with my having unexercised powers. For example, I would "have access to" worlds in which there are ten planets in the solar system. I would

[16]See van Inwagen, *Essay on Free Will,* 102–3, and "Reply to Christopher Hill," *Analysis* 52 (1992): 56–61.

[17]Van Inwagen, *Essay on Free Will,* 86–91.

[18]Ibid., 97.

also "have access to" worlds in which other people make choices different from the choices they actually make. But this doesn't indicate anything about what I can actually do or bring about, and so is irrelevant to my free will. The result of such a definition of the having-access-to relation, then, would be that the Consequence Argument would be valid, but it would not be an argument about free will. (Van Inwagen, of course, does not propose any definition of this kind; he just does not define the having-access-to relation at all.)

I will argue presently, then, that there is no way of specifying the set of worlds S has access to, such that S's having a choice about the fact that p is plausibly thought of as his having access to a world in which $\sim p$. For this purpose, I will exploit the interplay between two different examples. First example:

> Assume John and Sally both have free will, and that free choices, in general, are undetermined. John is able to give Sally $1. If he were to give Sally $1, she would then have a choice as to what to do with it: she could buy a bus ticket with it, or she could refrain from buying a bus ticket with it. John, however, chooses to keep his money.

Now, let

A = John gives Sally $1.
B = John gives Sally $1, and Sally buys a bus ticket with it.
C = John gives Sally $1, and Sally refrains from buying a bus ticket with it.

John's relation to B is the same as his relation to C—that is, he is no differently situated relative to Sally's possible choice to buy a bus ticket than he is relative to her possible choice to not buy a bus ticket (assume that John could not influence Sally's choice, once she received the dollar). John is in a position to give Sally the choice, but it is indeterminate which choice Sally would make if John were to do so: in some of the nearby possible worlds in which John gives her the dollar, she buys a bus ticket, and in some of the nearby worlds in which John gives her the dollar, she refrains. Thus, it would be arbitrary to say that John had access to the worlds in which B holds but not the ones in which C holds, or vice versa. John must have access either to both sets of worlds, or to neither. But we know John has access to some worlds in which

A holds—this follows from the fact that he has a choice about whether to give Sally $1. Every A-world is also either a B-world or a C-world. So John must have access to some worlds in which either B or C holds. And so he must have access, both to the worlds in which B holds, and to the worlds in which C holds.

This has the consequence that John has a choice about whether Sally buys a bus ticket (though Sally, herself, does not)[19]: In some of the worlds John has access to, she buys a ticket, and in some of the worlds he has access to, she doesn't. This is counterintuitive, since we have stipulated that John couldn't influence Sally's choice once she had been given the dollar. (It is also, of course, incompatible with the result of defining 'has a choice about' in terms of 'can render false', but that is what we want, since we have abandoned the 'can render false' definition.)

We can generate a more strongly counterintuitive result using a second example:

> Assume that both I and Mikhail Gorbachev have free will. At the present moment, I am deciding whether to drink a glass of milk or not. Shortly after I make this choice, Gorbachev will decide whether to take a walk or not. Neither of us, however, is able to exercise any influence on the other. I decide not to drink the glass of milk.

Now, let

D = I drink a glass of milk.
E = I drink a glass of milk, and Gorbachev takes a walk.
F = I drink a glass of milk, and Gorbachev refrains from taking a walk.

I have access to worlds in which D holds. Now, it seems that the same argument that showed that John had access to both the B-worlds and the C-worlds must also show that I have access to both the E-worlds and the F-worlds. Therefore, I have a choice about whether Gorbachev takes a walk or not (in some of the worlds I have access to, he takes a walk, and in some of the worlds I have

[19]Sally has no choice about whether John gives her the dollar. She has access to a world in which ~A (the actual world), but she lacks access to any worlds in which A holds; therefore, she lacks access to any world in which either B or C holds. Of course, if John *were* to give her the dollar, she *would then* have access to such worlds.

access to, he does not). But this is absurd—I clearly do not have any sort of control over Gorbachev. This shows that 'I have a choice about whether *p*' cannot be equated with 'In some of the possible worlds I have access to, *p,* and in some of the worlds I have access to, $\sim p$'.

But the following objection can be made to this last bit of reasoning. Unlike in the case of propositions B and C, there *is* a relevant distinction between E and F. Although E and F are both false, one of them has a second conjunct that is true. That is, either Gorbachev will take a walk in the actual world, or he won't. Let's stipulate that in the actual world,[20] Gorbachev will take a walk. In that case, this very fact can be taken as a reason for saying that I have access to E-worlds but not F-worlds: S has access to a world only if the choices made by people other than S in that world are the same as the choices made by people other than S in the actual world.[21]

It is in view of this sort of move that I introduced the first example. The first example shows that one must be allowed to "have access to" worlds in which other people make choices different from those made in the actual world—otherwise, John would have access neither to any B-worlds nor to any C-worlds, and so would not have access to any A-worlds. But John's relationship to the B-worlds in the first example, I claim, is relevantly similar to my relation to the F-worlds in the second example. John has no more control over Sally's choice than I have over Gorbachev's. The most that can be said, apropos of why John "has access to" worlds in which Sally buys a bus ticket, is that if John were to give Sally the dollar, she *might* buy a bus ticket. But it is also true that if I were to drink a glass of milk, Gorbachev *might* refrain from taking a walk. This is simply because it is already true that Gorbachev might or

[20]This locution sounds funny—one cannot stipulate what goes on in the actual world. In the context of this paragraph, 'the actual world' refers to the world that we are pretending is actual, when considering the Gorbachev example.

[21]Van Inwagen's remarks (*Essay on Free Will,* 89) suggest that he would want to say S only has access to worlds in which other people's choices are the same as in the actual world. But he does not specify what should be said about the situation in which one person's choices affect what choices others have.

might not take a walk, and my drinking the milk wouldn't alter the situation.

The possible-worlds interpretation of 'having a choice about the fact that *p*' abandons the requirement that my available action be *sufficient* for $\sim p$ in any sense (in no sense can John do something that is *sufficient* either for B or for C). It is this feature that enables the definition to avoid the counterexamples of section 2. But it is also this feature that makes it implausible as an interpretation of what it is to have a choice about something. For if I cannot do anything that is sufficient for $\sim p$, then intuitively, I do not have control over whether *p*. For example, John has no control over whether Sally buys a bus ticket or refrains from doing so, since he cannot do anything to guarantee either outcome. And it is implausible to hold that one "has a choice about" things which one cannot control.

4. The 'No Matter What' Interpretation

By strengthening the premises of an invalid argument, one can often make the argument valid, without sacrificing the plausibility of the premises. It is this strategy that I propose to take with the Consequence Argument. I propose to understand '$N_S p$' as follows:

$N_S p$ = No matter what S does, *p*

where

(No matter what S does, *p*) = (*p*, and for each action, A, that S can perform, if S were to perform A, it *would* still be the case that *p*).[22]

Thus, for example, '$N_I(2+2=4)$' appears to be true (where the subscript 'I' refers to me). To see this, consider the actions available to me at the moment. I can stand up; I can drink a glass of water; I can (as I am doing presently) just sit here typing on my keyboard; and so on.[23] Now, if I were to stand up, 2 + 2 would still

[22]McKay and Johnson ("A Reconsideration," 119) discuss this sort of interpretation with their proposed rule β4. In responding to the Frankfurt examples, van Inwagen ("Ability and Responsibility," *Philosophical Review* 87 (1978): 201–24, at 215–17) has argued that if, no matter what an agent had done, state S would have obtained, then the agent is not responsible for S. This suggests that he would be receptive to my proposed reinterpretation of the consequence argument.

[23]This, of course, is only true assuming I have free will. If I lack free

equal 4. If I were to drink a glass of water, 2 + 2 would still equal 4. If I were to just sit here typing these words, 2 + 2 would still equal 4. And a similar thing might be said for each action presently available to me. This is to say that no matter what I do, 2 + 2 = 4.

Notice that 'no matter what I do, *p*' is stronger than 'I have no choice about the fact that *p*,' as van Inwagen understands the latter. 'No matter what I do, *p*' requires that, for each act A that I can perform, the proposition that I perform A *be counterfactually sufficient for* p. But 'I have no choice about the fact that *p*', according to the definition of section 2, only requires that the proposition that I perform A *fail to suffice for* ~p. So I could be in a situation in which *p* is true, I can perform some act A, and if I were to perform A, *p* might or might not hold.[24] In that case, it is true that I have no choice about the fact that *p*, since I cannot render *p* false, but it is false that *no matter what I do, p*. To put this in terms of possible worlds: the truth of 'I have no choice about the fact that *p*' requires that *p* be true, and (for every act A that I can perform) in some of the nearby worlds in which I perform A, *p* continue to hold. But the truth of 'no matter what I do, *p*' requires that *p* be true, and (for every act A that I can perform) in *all* of the nearby worlds in which I perform A, *p* continue to hold.

Now let's see how the present definition of 'N*p*' avoids the counterexample to β and β* presented in section 2. In that case, S can activate the device to shoot an R-particle into the basket. It is not the case that if S were to do so, no R-particle would enter the left half of the basket; rather, if S were to do so, an R-particle *might* enter the left half of the basket. Therefore, N_SA is false. Similarly, N_SB is false. Thus, no counterexample can be generated.

Moreover, it is easy to see that rules α, β, α*, and β* must be valid for N_Sp as defined. Assume that for each act A that S can perform, in all the nearby worlds in which S performs A, *p* holds. Assume also that for each act A that S can perform, in all the nearby worlds in which S performs A, *q* holds. Then, obviously, for each act A that S can perform, in all the nearby worlds in which S performs A, both *p* and *q* hold. Therefore, rule β* is valid. Rule α* is also valid: assume that for every act A that S can perform, in

---

will, then the only action I can perform is that of sitting here typing just these words on my keyboard.

[24]Assume A is the only relevant non-actual action available to me.

all the nearby worlds in which S performs A, p holds. Assume also that in every possible world in which p holds, q holds. Then, obviously, for every act A that S can perform, in all the nearby worlds in which S performs A, q holds. (Similar considerations apply to rules α and β.) Starting from the assumption of determinism, then, we may validly argue:

1. $(P_0 \ \& \ L) \rightarrow P$	assumption
2. $N_S P_0$	premise
3. $N_S L$	premise
4. $N_S(P_0 \ \& \ L)$	2, 3; rule β*
5. $N_S P$	1, 4; rule α*

where S is any arbitrarily chosen person.

There is no danger that, in redefining '$N_S P$', we have made the argument irrelevant to free will. It is true that step 5 no longer means that S has no choice about the fact that P. Rather, step 5 says something logically stronger than that S has no choice about the fact that P. It says that no matter what S does, P would still hold, and this clearly entails that S lacks free will with respect to the fact that P.

5. The Plausibility of the Consequence Argument

In order for the above version of the Consequence Argument to prove that the existence of free will is incompatible with determinism, three conditions must hold. First, rules α* and β* must be valid. Second, the universal closure of $N_S p$ must imply the absence of free will. Third, premises $N_S P_0$ and $N_S L$ must be true, for any arbitrarily chosen person, S. As we have just seen, the first two conditions clearly hold when '$N_S p$' is understood in the manner explained in section 4. It remains only to consider whether the premises of the Consequence Argument are still plausible, under this strengthened interpretation of '$N_S p$'.

I think the following principle is highly plausible: if p is any fact about what went on in the past, then $N_S p$. For example, consider the proposition that Napoleon lost the Battle of Waterloo in 1815. And consider once again the range of actions available to me at the moment. I can stand up; I can drink a glass of water; I can sit

here typing; and so on. Well, Napoleon lost the Battle of Waterloo in 1815, and if I were to stand up now, it would still be the case that Napoleon lost the Battle of Waterloo in 1815. If I were to drink a glass of water now, it would still be the case that Napoleon lost the Battle of Waterloo in 1815. If I were to sit here typing, it would still be the case that Napoleon lost the Battle of Waterloo in 1815. And so on. No matter what I do, Napoleon lost the Battle of Waterloo in 1815. This much seems entirely obvious. Nor is there anything special about Napoleon's defeat at Waterloo—the same sort of intuition holds for anything that took place in the past.

It is scarcely less obvious that, if p is any law of nature, then $N_S p$. For example, take the law of the conservation of matter,[25] and consider the same range of actions open to me at present. Matter can be neither created nor destroyed, and if I were to stand up right now, it would still be the case that matter can be neither created nor destroyed. If I were to drink a glass of water, it would still be the case that matter can be neither created nor destroyed. If I were to sit here typing, it would still be the case that matter can be neither created nor destroyed. Nor is there anything special about this particular law—a similar sort of intuition holds for any law of nature.[26]

At least one philosopher, however, rejects these intuitions. David Lewis claims that, if A is any action I do not perform, if I were to perform A, some actual law of nature would be false *and* the past would be different from the actual past.[27] In his terminology, my performing A would render a law false in the weak sense. This is

---

[25]Strictly speaking, of course, we should talk about the conservation of mass/energy, but this qualification is irrelevant to our present concerns.

[26]By 'law of nature' I refer to physical laws, such as the law of gravity, the conservation of momentum, F=ma, and so on. The intuition in question does not clearly hold for, say, laws of psychology (if there are any such things). Van Inwagen (*Essay on Free Will,* 61–62) makes a similar stipulation.

[27]Lewis, "Are We Free to Break the Laws?" 116–17. More precisely, Lewis claims that this is true *if* determinism is true, and that determinism *could* be true while at the same time we had free will. John Turk Saunders argues, anticipating Lewis, that sometimes a person can act in such a way that, if he did so, the past would have been different than it actually was, even though no one can *cause* anything to happen in the past ("The Temptations of Powerlessness," *American Philosophical Quarterly* 5 (1968): 100–8). But see also John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control* (Oxford: Blackwell, 1994), 81–82, which responds (albeit inconclusively) to Saunders's argument.

because, if I performed act A, it would have to have been because a divergence miracle occurred shortly before my performance of A.[28]

Now, it is not entirely implausible that this is true, if determinism is true. But what *is* entirely implausible is that, in spite of this, I might be perfectly free to do A. If, in order for me to do A, something would have to have happened five minutes ago (or a year ago, or five seconds ago, or whatever) that did not in fact occur, then I cannot now do A.[29] If someone disagrees with this, the best that I can do is appeal to instances of this general principle that seem obviously correct:

(a) Suppose that, in order for us to avoid a serious degradation of the earth's ozone layer, we would have to have instituted policies to curb the use of chlorofluorocarbons twenty years ago. Suppose that, in fact, we did nothing of the kind. Then it obviously follows that it is not, now, within our power to avoid a serious degradation of the ozone layer.

(b) Suppose that you are a doctor in the emergency room of a hospital, where a heart attack victim has just been brought in. Suppose you know, as a matter of the laws of biochemistry and the physiology of the human body, that in order for a heart attack patient to be revived, CPR must be administered within three minutes of the time of the heart attack. Suppose you also know that for this patient, four minutes have already elapsed, during which the patient did not receive CPR. You would obviously be correct to conclude that, at this point, you cannot revive the patient.

I do not see how either the importance of the required past event or how long ago it would have to have happened could make a

[28]See David Lewis, "Counterfactual Dependence and Time's Arrow," *Noûs* 13 (1979): 455–76, for an explanation of "divergence miracles." A divergence miracle is a small violation of (what in the actual world is) a law of nature—the minimal violation required in order to make true the antecedent of some counterfactual conditional we are considering.

[29]Fischer sympathetically discusses essentially this principle, which he calls "the fixity of the past" (see especially 78–85). Fischer's formulation is "no person can act in such a way that some fact about the past would not have been a fact" (9). Carl Ginet also defends a related "principle of the fixity of the given past" (*On Action* (Cambridge: Cambridge University Press, 1990), 102–10).

difference to the argument, so I conclude that if, in order for me to do A, a small miracle would have to have happened a short time ago, whereas no such miracle in fact happened, then I cannot now do A.

Now a worry may arise that I have been unfair to Lewis. Lewis's claim, with respect to the imagined case where I am determined not to do A, was not that in order for me to do A, a certain divergence miracle *would have to have* happened a short time previously. Lewis's claim was rather that, if I were to do A, a divergence miracle *would have* happened a short time previously. The latter seems to be a weaker statement: for perhaps there are many different ways that I *could* have come to do A, but only one way that I *would* have. The way that I would have come to do A if I did is the way that involves the least deviation from the actual world, but (depending on how 'could' is read) the ways that I could have come to do A may be taken to include a great many more ways.

More generally: my principle is that if, in order for me to do A, B would have to have happened at some previous time, whereas in fact B did not happen, then I cannot do A. A soft determinist might deny that his position conflicts with this principle, because although, in order for me to do A, the past would have to have been different in *some* way, there is no event-type B such that, in order for me to do A, B in particular would have to have happened. Rather, there are many different past events that would do.

But this response gains no intuitive advantage, for I might just as well state the principle as follows: If, in order for me to do A, either B_1 or B_2 or . . . or B_n would have to have happened, whereas in fact none of $\{B_1 \ldots B_n\}$ happened, then I cannot do A. This principle can be supported by similar examples. Modify example (a) as follows: assume that in order to avoid the degradation of the ozone layer, we would have to have implemented at least one of twelve different policies, and done so some time between ten and twenty years ago. Assume that in fact, we implemented none of them. Again, it obviously follows that it is not now in our power to prevent the degradation of the ozone layer.

Similarly, suppose that, in example (b), there are three different ways of reviving a heart attack victim. Suppose that, in order to save the victim, either of the first two would have to have been applied within three minutes of the heart attack, or the third applied within three minutes and thirty seconds. Suppose that in fact,

none of them was applied within the required time interval. Again, one could obviously infer that, at this point, the patient cannot be revived.

The general principle is that if, in order for one to do anything other than what one actually does, the past would have to have differed from the actual past, then one does not have free will. Lewis has not refuted this contention. He has distinguished between a strong and a weak sense of 'can render false', but to no effect, for he has said nothing to support the implausible assertion that one can render false a law of nature or a statement about the past, even in the weak sense. Thus, the Consequence Argument remains a very powerful argument for incompatibilism.

University of Colorado at Boulder