The Consequence of the Consequence Argument[*]

Marco Hausmann

Abstract

The aim of my paper is to compare three alternative formal reconstructions of van Inwagen's famous argument for incompatibilism. In the first part of my paper, I examine van Inwagen's own reconstruction within a propositional modal logic. I point out that, due to the expressive limitations of his propositional modal logic, van Inwagen is unable to argue directly (that is, within his formal framework) for incompatibilism. In the second part of my paper, I suggest to reconstruct van Inwagen's argument within a first-order predicate logic. I show, however, that even though this reconstruction is not susceptible to the same objection, this reconstruction can be shown to be inconsistent (given van Inwagen's own assumptions). At the end of my paper, I suggest to reconstruct van Inwagen's argument within a quantified counterfactual logic with propositional quantifiers. I show that within this formal framework van Inwagen would not only be able to argue directly for incompatibilism, he would also be able to argue for crucial assumptions of his argument.

Keywords: *free will, determinism, consequence argument, counterfactual logic*

1 Introduction

Incompatibilism, as I understand it, is the view that there is no free will if determinism is true. Peter van Inwagen has developed a much discussed argument for incompatibilism. His argument runs as follows:

*This paper won the *SOPhiA 2019 Best Paper Award* at the *Salzburg Conference for Young Analytic Philosophy*, September 18–20, 2019.

Kriterion – Journal of Philosophy, 2020, 34(4): 45–70.
http://www.kriterion-journal-of-philosophy.org

> „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. I shall call this argument the Consequence Argument" [15, p. 56].

According to van Inwagen's argument, if determinism is true, our acts are the consequences of the laws of nature and events in the remote past. However, neither the laws of nature nor events in the remote past are up to us. It follows that, if determinism is true, our acts are not up to us.

In my view, one of van Inwagen's major achievements is his attempt to reconstruct his argument within a formal framework. In the first part of my paper, I briefly recapitulate van Inwagen's attempt to reconstruct his argument within a propositional modal logic. It will become evident, however, that van Inwagen's attempt has a shortcoming: the conclusion of van Inwagen's formal argument is *not* the conclusion of van Inwagen's informal argument. The conclusion of van Inwagen's formal argument is not the conclusion that our acts are not up to us if determinism is true (let alone the conclusion that there is no free will if determinism is true). At the end of the first part, I explain why this might put some pressure on defenders of van Inwagen's argument. In the second part of my paper, I suggest an alternative reconstruction. I reconstruct van Inwagen's argument within a first-order predicate logic. As it turns out, the conclusion of this formal argument *is* the conclusion of van Inwagen's informal argument. However, as I point out at the end of the second part, there is reason to think that this version of van Inwagen's argument makes use of a formal language that leads into paradox. In the third part of my paper, I suggest an alternative and, in my view, better reconstruction. I reconstruct van Inwagen's argument within a quantified counterfactual logic with propositional quantifiers. I show that the conclusion of van Inwagen's argument within a quantified counterfactual logic is the conclusion that there is no free will if determinism is true. Besides that, I show that within a quantified counterfactual logic van Inwagen would be able to argue for crucial assumptions of his argument. Thus, a quantified counterfactual logic appears to be better suited for the task of an adequate formal representation of van Inwagen's famous argument.

2 *The Consequence Argument within Propositional Modal Logic*

As I said, van Inwagen's first suggestion was to reconstruct his argument within a propositional modal logic [15, p. 93–105]. To this end, he introduced a modal operator 'N' (where 'Np' stands for 'p and no one has, or ever had, any choice about whether p').[1] He conjectured that, given this definition of 'N', the following schemata would turn out to be valid:

Rule Alpha:[2] $\Box p \vdash Np$

Rule Beta: $Np,\ N(p \rightarrow q) \vdash Nq$

He then defined determinism as the conjunction of two theses:

> "For every instant of time, there is a proposition that expresses the state of the world at that instant;
> If p and q are any propositions that express the state of the world at some instants, then the conjunction of p with the laws of nature entails q" [15, p. 65].

The main idea of his formal argument is, roughly, the following: Take an arbitrary truth. If determinism is true, the conjunction of a past state of the world with the laws of nature entails that truth. But nobody has a choice about whether the conjunction of a past state of the world with the laws of nature is true. Therefore, nobody has a choice about that truth.

In fact, if we let 'P_0' be a proposition that expresses the state of the world in the remote past (at a time long before the existence of the first human being), if we let 'P' be an arbitrary truth, and if we let 'L' be the conjunction of all laws of nature, then we can argue as follows:

(1) $\Box((p_0 \& l) \rightarrow p)$ — follows from determinism

(2) $\Box(p_0 \rightarrow (l \rightarrow p))$ — follows from (1)

(3) $N(p_0 \rightarrow (l \rightarrow p))$ — follows from (2) and rule alpha

(4) Np_0 — premise

(5) $N(l \rightarrow p)$ — follows from (3), (4) and rule beta

(6) Nl — premise

(7) Np follows from (5), (6) and rule beta

The main idea of van Inwagen's formal argument, therefore, is that we can take an arbitrary truth and that we can show that nobody has a choice about that truth if we suppose that determinism is true. Thus, while van Inwagen's informal argument purports to show that our acts are not up to us if determinism is true, van Inwagen's formal argument purports to show that nobody has a choice about an arbitrary truth if determinism is true. The conclusion of van Inwagen's informal argument, therefore, is *not* the conclusion of van Inwagen's formal argument. Strictly speaking, van Inwagen's formal argument is not an argument for incompatibilism anymore.

It is not entirely clear whether this is a severe drawback. For it seems, at least intuitively, that the conclusion of van Inwagen's formal argument *entails* the conclusion of van Inwagen's informal argument. In my view, however, there are at least two reasons why we should look for a formal framework in which we can argue *directly* for incompatibilism.

First, in arguing for incompatibilism within a formal framework, we would have to develop a rigorous formal proof and we would, thereby, make transparent what one would have to deny in order to deny that van Inwagen's argument is an argument for incompatibilism. This might shed light on van Inwagen's background assumptions and help participants of the debate to "measure the price" of rejecting van Inwagen's claim that his argument is an argument for incompatibilism. As I will show at the end of my paper, this might enable critics of van Inwagen's argument to get clear about which premises of van Inwagen's argument, if any, ought to be denied.

Second, this might remove doubts about van Inwagen's claim that his argument is an argument for incompatibilism – doubts one might otherwise have. In order to understand why one might have doubts about van Inwagen's claim, we have to take note of the fact that van Inwagen had to revise his original argument and that, with respect to his revised argument, van Inwagen's conclusion has come under attack. More specific, after Thomas McKay and David Johnson [9] had been able to present a convincing counterexample against van Inwagen's rule beta, van Inwagen had to revise his original interpretation of 'N'. He came to suggest a modified interpretation of 'N' according to which 'Np' stands for 'p and every region of logical space to which anyone has, or ever had, exact access is a subregion of p' [16, p. 8–10]. Thus, the conclusion of van Inwagen's revised argument is that we can show with respect to an arbitrary truth that every region of logical space to which anyone has, or ever had, exact access is a subregion of that truth. This is not the place to go into the details of van Inwagen's definition. What matters is that, contrary to what one would have expected, Lynne Rudder Baker [1] appears to have been able to show that van Inwagen's conclusion "immediately follows, whether determinism is true or not" [1, p. 16]. Thus, if van Inwagen's assumption were true (that from his conclusion it follows that our acts are not up to us and that there is no free will), it would immediately follow, regardless of whether determinism is true or not, that our acts are not up to us and that there is no free will. To many, this conclusion is unacceptable. For this reason, Baker concludes: "Van Inwagen may interpret my result as implying that free will is incompatible with both determinism and indeterminism, and hence is incoherent. [...] I would not draw such a conclusion. [...] Perhaps the conclusion of the Consequence Argument has nothing to do with free will, properly conceived" [1, p. 21, fn. 13]. Thus, at least insofar as there is reason to think that free will is *not* incoherent, there is reason to think, contrary to what one would have expected, that van Inwagen's assumption (that from his conclusion it follows that our acts are not up to us and that there is no free will) is *false*.

Peter van Inwagen has, therefore, suggested a different interpretation of 'N' according to which 'Np' stands for 'p and no human being is or ever has been able to act in such a way that, if he or she did act that way, it might be or might have been false that p' [18, p. 214]. In his view, once we take this interpretation of 'N', Baker's "argument does not apply" [19, p. 12, fn. 21]. However, given Baker's surprising (and, in a sense, counterintuitive) result, one might still have doubts about van Inwagen's conclusion (one might still suspect that, one day, a *different* argument with the same result might be discovered). Thus, one might still have doubts about van Inwagen's conclusion – at least as long as we do not remove that doubts by arguing *directly* for incompatibilism within a formal framework.

To be sure, nothing I have said amounts to a serious objection against van Inwagen's argument. After all, van Inwagen's argument, like every other argument, has to stop somewhere. What I have said simply suggests that it would be desirable to reconstruct van Inwagen's argument within a formal framework in which we can argue directly for incompatibilism. In the reminder of this paper, I try to reconstruct van Inwagen's argument within formal frameworks in which we can argue directly for incompatibilism.

3 *The Consequence Argument within First-Order Predicate Logic*

In a recent paper, van Inwagen actually seems to suggest a formal framework in which we can argue directly for incompatibilism. He starts with three definitions:

> "Say that a proposition p is *untouchable* just in the case that: p is true and no human being is or ever has been able to act in such a way that, if he or she did act that way, p might be or might have been false.
> Say that, for any time (=instant of time) t, a *t-state proposition* is a proposition that gives a complete description of the state of the world at t [. . .].
> Say that a human being is *multiply able* if there was an occasion on which he or she was trying to decide which of two (or more) incompatible alternative courses of action (e.g. lying and telling the truth) [to perform] and was at some point in the course of those deliberations able to perform each of them" [18, p. 214].

Thus, van Inwagen no longer speaks of nobody having a choice about truths. Instead, he speaks of truths being untouchable. His definition of being untouchable is meant to avoid the counterexample against rule beta that has been developed by McKay and Johnson. Further, van Inwagen no longer speaks of acts being up to us. Instead, he speaks of human beings being multiply able. He goes on to argue as follows:

> "The Consequence Argument has the following seven premises:
>
> 1. Every necessary truth is untouchable
> 2. The conjunction of all laws of nature (= &L) is untouchable
> 3. For every time t, if there were as yet no human beings at t, the t-state proposition is untouchable
> 4. There is a time t at which there were as yet no human beings
> 5. If p is untouchable and if the conditional whose antecedent is p and whose consequent is q is untouchable, then q is untouchable

6. If the world is deterministic, then, for every true proposition p, and every time t, the conditional whose antecedent is &L and whose consequent is (the conditional whose antecedent is the t-proposition and whose consequent is p) is a necessary truth
7. If any human being is or ever has been multiply able, then some true proposition is not untouchable

> And its conclusion is:
> If the world is deterministic, no human being is or ever has been multiply able.
> The demonstration that this argument is logically valid is left as an exercise for the reader. (Textbook quantifier logic suffices)" [18, p. 214–15].

In my view, even though van Inwagen offers no formal translation of his argument, his claim that textbook quantifier logic suffices for a demonstration "that this argument is logically valid" strongly suggests a reconstruction of his argument within first-order predicate logic.

First of all, however, it might prove useful to simplify van Inwagen's argument. To this end, let us take a look at the fourth and the sixth premise:

> "4. There is a time t at which there were as yet no human beings
> [...]
> 6. If the world is deterministic, then, for every true proposition p, and every time t, the conditional whose antecedent is &L and whose consequent is (the conditional whose antecedent is the t-proposition and whose consequent is p) is a necessary truth" [18, p. 214–15].

Let us say that a state of the world is a *prehuman* state of the world just in case that, at that state of the world, "there were as yet no human beings". The fourth premise suggests, accordingly, that there is a prehuman state of the world. The fourth and the sixth premise taken together entail, therefore, that, if the world is deterministic, then a conjunction of the conjunction of all laws of nature with a description of a prehuman state of the world entails all truths.[3] I suggest, therefore, to take the fourth and the sixth premise of van Inwagen's argument together:

> (1) The thesis of determinism is true $\rightarrow$ $\exists x$ (x is a conjunction of the conjunction of all laws of nature with a complete

> description of a prehuman state of the world & $\forall y$ (y is true $\rightarrow x$ entails y))

Take a look, further, at the first and the fifth premise of van Inwagen's argument:

> "1. Every necessary truth is untouchable
> [. . .]
> 5. If p is untouchable and if the conditional whose antecedent is p and whose consequent is q is untouchable, then q is untouchable" [18, p. 214–15].

Note that van Inwagen simply *assumes* that every necessary truth is untouchable. He is unable to *show* that every necessary truth is untouchable because, in order to do that, he would have to define the notion of a necessary truth, he would have to formally represent his definition of untouchability and he would have to argue that, given his definitions, every necessary truth is untouchable. However, this is not a straightforward task within a first-order predicate logic.

Be that as it may, the first and the fifth premise taken together entail that if P is untouchable and if P entails Q, then Q is untouchable.[4] I suggest, therefore, to take the first and the fifth premise of van Inwagen's argument together:

> (2) $\forall x \forall y$((x is untouchable & x entails y) $\rightarrow y$ is untouchable)

Note, further, that the first and the fifth premise taken together entail that if P is untouchable and if Q is untouchable, then the conjunction of P and Q is untouchable.[5] This result enables us to take the second and the third premise of van Inwagen's argument together:

> "2. The conjunction of all laws of nature (= &L) is untouchable
> 3. For every time t, if there were as yet no human beings at t, the t-state proposition is untouchable" [18, p. 214].

For this suggests that if the conjunction of all laws of nature is untouchable and if every complete description of a prehuman state of the world is untouchable, then every conjunction of the conjunction of all laws of nature with a complete description of a prehuman state of the world is untouchable as well:

> (3) $\forall x$(x is a conjunction of the conjunction of all laws of nature with a complete description of a prehuman state of the world $\rightarrow x$ is untouchable)

All we need to do now is to formally represent the seventh premise of van Inwagen's argument:

> "7. If any human being is or ever has been multiply able, then some true proposition is not untouchable" [18, p. 215].

As it turns out (a few minor details aside), van Inwagen regards the thesis that human beings are multiply able as equivalent to the thesis that our acts are up to us and, further, as equivalent to the thesis that there is free will. For in another recent paper, he writes:

> "Let us say that it is at a certain moment *up to one whether* one will do A or do B if one is then faced with a choice between doing A and doing B and one is then able to do A and is then able to do B [...]" [21, p. 166].

In yet another recent paper, he writes:

> "The free-will thesis is the thesis that we are sometimes in the following position with respect to a contemplated future act: we simultaneously have both the following abilities: the ability to perform that act and the ability to refrain from performing that act" [20, p. 151].

Thus, van Inwagen regards the thesis that human beings are multiply able as equivalent to the thesis that our acts are up to us and, further, as equivalent to the thesis that there is free will. Thus, instead of saying that if any human being is multiply able then some truth is not untouchable, we may equally say that if there is free will, then some truth is not untouchable, or, alternatively, that if all truths are untouchable, then there is no free will:

(4) $\forall x(x \text{ is true} \rightarrow x \text{ is untouchable}) \rightarrow \sim \exists x(x \text{ is free will})$

Note, again, that van Inwagen has to *assume* that there is no free will if all truths are untouchable. He is unable to *show* that there is no free will if all truths are untouchable. In order to do that, he would have to formally represent his definition of untouchability as well as his definition of being multiply able and he would have to argue that, given his definitions, nobody is multiply able if all truths are untouchable. Again, this is not a straightforward task within a first-order-predicate logic. Be that as it may, here finally is my reconstruction of van Inwagen's argument within first-order predicate logic:

(1) The thesis of determinism is true $\rightarrow \exists x$ (x is a conjunction of the conjunction of all laws of nature with a complete description of a prehuman state of the world & $\forall y$ (y is true $\rightarrow x$ entails y))

(2) $\forall x\ \forall y$ ((x is untouchable & x entails y) $\rightarrow y$ is untouchable)

(3) $\forall x$ (x is a conjunction of the conjunction of all laws of nature with a complete description of a prehuman state of the world $\rightarrow x$ is untouchable)

(4) $\forall x$ (x is true $\rightarrow x$ is untouchable) $\rightarrow \sim \exists x$ (x is free will)

The conclusion of van Inwagen's argument is the conclusion that there is no free will if the thesis of determinism is true:

(5) The thesis of determinism is true $\rightarrow \sim \exists x$ (x is free will)

This argument is obviously logically valid. Thus, if we reconstruct van Inwagen's argument within first-order predicate logic, we can reconstruct van Inwagen's argument as arguing directly for incompatibilism.

To my mind, the most important difference between this reconstruction of van Inwagen's argument and van Inwagen's original reconstruction within a propositional modal logic is that this reconstruction represents modalities by means of predicates and not by means of operators. Predicates are expressions that applied to a singular term yield a sentence. Thus, if we represent untouchability by means of a predicate (such as 'is untouchable'), we can apply this predicate to a singular term (such as 'the conjunction of all laws of nature') in order to yield a sentence:

(1) The conjunction of all laws of nature is untouchable

What matters (for our purposes) is that, if we represent untouchability as a predicate, we can not only express simple sentences (such as 'the conjunction of all laws of nature is untouchable'), we can also express sentences that involve quantification over all truths. In particular, we can express the sentence that all truths are untouchable:

(2) $\forall x$ (x is true $\rightarrow x$ is untouchable)

Operators, on the other hand, are expressions that applied to a sentence yield another sentence. Thus, if we represent untouchability by means of an operator (such as 'it is untouchable that'), we can apply this operator to a sentence (such as '$2 + 2 = 4$') in order to yield another sentence:

(3) It is untouchable that $2 + 2 = 4$

However, unlike the case where we represent untouchability by means of a predicate, we *cannot* express the sentence that all truths are untouchable (without using propositional quantifiers). This obviously speaks in favor of a reconstruction of van Inwagen's argument within first-order predicate logic. For within first-order predicate logic, we can argue directly for the conclusion that all truths are untouchable if determinism is true and, therefore, for the conclusion that there is no free will if determinism is true. By comparison, in a propositional modal logic we can at best show that it is untouchable that p if determinism is true (where 'p' is an abbreviation for an arbitrary truth). We cannot argue directly for the conclusion that all truths are untouchable if determinism is true and, therefore, for the conclusion that there is no free will if determinism is true.

However, as is well known, Richard Montague [10] has shown that many formal languages that represent *necessity* by means of a predicate are inconsistent. In what follows, I will extend his result in order to show that many formal languages that represent *untouchability* by means of a predicate are inconsistent as well. In a nutshell, Montague's argument is this:[6] It is plausible to maintain that only truths are necessary and that logical truths are necessary truths. In a formal language in which necessity is represented by means of a predicate, the following schemata would therefore have to be valid:

(***Factivity of Necessity***) If $\ulcorner p \urcorner$ is necessary then p

(***Necessity of Logic***) If $\vdash p$ then $\ulcorner p \urcorner$ is necessary

However, let $\ulcorner p \urcorner$ abbreviate a sentence that is true if and only if it is not necessary:

(***Self-Referential Sentence***) $p \leftrightarrow \sim$ ($\ulcorner p \urcorner$ is necessary)

Intuitively, $\ulcorner p \urcorner$ says of itself that it is not necessary. As is well known, we can prove that there is such a sentence if we name all expressions of our formal language by Gödel numbering. It is not difficult to show, therefore, that such a formal language is inconsistent. For such a formal language entails that $\ulcorner p \urcorner$ is necessary and not necessary. To see this, suppose that $\ulcorner p \urcorner$ is necessary:

(1) $\ulcorner p \urcorner$ is necessary — assumption

It follows that p (given that only truths are necessary):

(2) p — (1), Factivity of Necessity

Therefore, $\ulcorner p \urcorner$ is not necessary (given that p if and only if $\ulcorner p \urcorner$ is not necessary):

(3) $\sim \ulcorner p \urcorner$ is necessary (2), Self-Referential Sentence

Thus, the assumption that $\ulcorner p \urcorner$ is necessary leads to a contradiction. It follows that $\ulcorner p \urcorner$ is not necessary:

(4) $\sim \ulcorner p \urcorner$ is necessary (1)–(3)

However, $\ulcorner p \urcorner$ is a sentence that is true if and only if it is not necessary. It follows that p:

(5) p (4), Self-Referential Sentence

It follows, further, that $\ulcorner p \urcorner$ is a logical truth (for we have been able to prove that $\ulcorner p \urcorner$ is true). We can conclude that $\ulcorner p \urcorner$ is necessary:

(6) $\ulcorner p \urcorner$ is necessary (1)–(5), Necessity of Logic

Hence, $\ulcorner p \urcorner$ is necessary and not necessary. This result suggests that formal languages that represent necessity by means of a predicate are inconsistent (given fairly innocuous assumptions about necessity). And it is not difficult to see that Montague's result can be extended to formal languages that represent untouchability by means of a predicate. For recall that, according to van Inwagen, "a proposition p is *untouchable* just in the case that: p is true and no human being is or ever has been able to act in such a way that, if he or she did act that way, p might be or might have been false" [18, p. 214]. Thus, according to van Inwagen, only truths are untouchable. Recall, further, that, according to van Inwagen, every "necessary truth is untouchable" [18, p. 214]. In a formal language in which untouchability is represented by means of a predicate, the following schemata would therefore have to be valid:

(***Factivity of Untouchability***) If $\ulcorner p \urcorner$ is untouchable then p

(***Untouchability of Necessity***) If $\ulcorner p \urcorner$ is necessary then $\ulcorner p \urcorner$ is untouchable

Further, in a formal language in which untouchability is represented by means of a predicate, logical truths would still have to be necessary truths, that is, the following schemata would still have to be valid:

(***Necessity of Logic***) If $\vdash p$ then $\ulcorner p \urcorner$ is necessary

However, let $\ulcorner p \urcorner$ abbreviate a sentence that is true if and only if it is not untouchable:

(Self-Referential Sentence) $p \leftrightarrow \sim (\ulcorner p \urcorner$ is untouchable$)$

Intuitively, $\ulcorner p \urcorner$ says of itself that it is not untouchable. It is not difficult to show that such a formal language is inconsistent as well. For such a formal language entails that $\ulcorner p \urcorner$ is untouchable and not untouchable. To see this, suppose that $\ulcorner p \urcorner$ is untouchable:

(1) $\ulcorner p \urcorner$ is untouchable — assumption

It obviously follows that p (given that only truths are untouchable):

(2) p — (1), Factivity of Untouchability

It follows, however, that $\ulcorner p \urcorner$ is not untouchable (given that p if and only if $\ulcorner p \urcorner$ is not untouchable):

(3) $\sim \ulcorner p \urcorner$ is untouchable — (2), Self-Referential Sentence

Thus, the assumption that $\ulcorner p \urcorner$ is untouchable leads to a contradiction. It follows that $\ulcorner p \urcorner$ is not untouchable:

(4) $\sim \ulcorner p \urcorner$ is untouchable — (1)–(3)

However, $\ulcorner p \urcorner$ is a sentence that is true if and only if it is not untouchable. It follows that p:

(5) p — (4), Self-Referential Sentence

It follows, further, that $\ulcorner p \urcorner$ is a logical truth (for we have been able to prove that $\ulcorner p \urcorner$ is true). It follows that $\ulcorner p \urcorner$ is necessary:

(6) $\ulcorner p \urcorner$ is necessary — (1)–(5), Necessity of Logic

And given that every necessary truth is untouchable, $\ulcorner p \urcorner$ is untouchable as well:

(7) $\ulcorner p \urcorner$ is untouchable — (6), Necessity of Untouchability

Thus, we have been able to prove that $\ulcorner p \urcorner$ is untouchable and not untouchable (given van Inwagen's assumptions about untouchability). I conclude that Montague's result does not only apply to formal languages that represent necessity by means of a predicate but also to formal languages that represent untouchability by means of a predicate.

Let us take a step back: incompatibilism is the view that there is no free will if determinism is true. One of van Inwagen's major achievements is his attempt to develop an argument for incompatibilism within a propositional modal logic. However, due to the expressive limitations of his propositional modal logic, van Inwagen is unable to argue directly for incompatibilism. He is unable to argue for the thesis that all truths are untouchable if determinism is true (and, therefore, for the thesis that there is no free will if determinism is true). For this reason, one might prefer to reconstruct van Inwagen's argument within a first-order predicate logic. However, in order to do that, one would have to represent untouchability by means of a predicate and, as we have seen, a formal language that represents untouchability by means of a predicate can be shown to be inconsistent (given van Inwagen's assumptions about untouchability).

One might, of course, try to restrict the schemata that give rise to this inconsistency.[7] One might, for example, restrict the schemata to *grounded truths* (and maintain that a truth is not grounded if it says of itself that it is not untouchable). However, the task of drawing a line between grounded truths and ungrounded truths turns out to be a very complex matter.[8] I fear, therefore, that van Inwagen's argument would lose much of its intuitive appeal if his pretty straightforward principles were replaced with principles that rely on a complex distinction such as the distinction between grounded and ungrounded truths. For this reason, I prefer to pursue a different strategy. I prefer to reconstruct van Inwagen's argument within a quantified counterfactual logic with propositional quantifiers. As will soon emerge, this is going to improve van Inwagen's argument in several respects. For within a quantified counterfactual logic we are not only able to argue directly for incompatibilism, we are also able to argue for important assumptions of van Inwagen's argument.

4 The Consequence Argument within Quantified Counterfactual Logic

In what follows, I am going to reconstruct van Inwagen's argument within a quantified counterfactual logic. To this end, I will assume that '$p \mathrel{\Box\!\!\rightarrow} q$' is an abbreviation for 'if it were the case that p, then it *would* be the case that q' and I will assume that '$p \mathrel{\Diamond\!\!\rightarrow} q$' is an abbreviation for 'if it were the case that p, then it *might* be the case that q'. Further, I will follow David Lewis [8, p. 2] in assuming that a sentence of the form 'if it were the case that p, then it *might* be the case that q' is equivalent to a sentence of the form 'it is *not* the case that if it were the case that p, then it *would not* be the case that q':

(*Interdefinability*) $p \Diamond\!\!\rightarrow q \leftrightarrow \sim (p \Box\!\!\rightarrow \sim q)$

Besides that, I will work with the following axioms that constitute a very weak subsystem of Lewis's preferred counterfactual logic:[9]

(*Tautology*) If p is a truth-functional tautology then $\vdash p$

(*Vacuity*) $(\sim p \Box\!\!\rightarrow p) \rightarrow (q \Box\!\!\rightarrow p)$

(*Weak Centering*) $(p \Box\!\!\rightarrow q) \rightarrow (p \rightarrow q)$

(*Equivalence*) If $\vdash (p \leftrightarrow q)$ then $\vdash (p \Box\!\!\rightarrow r) \leftrightarrow (q \Box\!\!\rightarrow r)$

Suppose that we add objectual as well as propositional quantifiers to our formal language (with appropriate elimination and introduction rules). All we have to do then (if we want to reconstruct van Inwagen's argument within this formal framework), is to define the notion of necessity and the notion of untouchability respectively.

First of all, following a suggestion of Lewis [8, p. 22], let us define the notion of necessity as follows:[10]

(*Necessity*) $\Box p \leftrightarrow (\sim p \Box\!\!\rightarrow p)$

According to Lewis [8, p. 22], this definition is meant to capture the intuition that it is necessary that p just in case that it would be the case, no matter what, that p. For if it would be the case, no matter what, that p, then it would still be the case that p, even if it were false that p (and vice versa).

Recall, further, that, according to van Inwagen, "a proposition p is *untouchable* just in the case that: p is true and no human being is or ever has been able to act in such a way that, if he or she did act that way, p might be or might have been false" [18, p. 214]. Following a suggestion of Huemer [4] and Pruss [12], this definition can be formally represented as follows (if we let 'Up' stand for 'it is an untouchable truth that p'):

(*Can't Touch I*)

$$Up \leftrightarrow (p \;\&\; \sim \exists x \exists y (x \text{ is able to do } y \;\&\; (x \text{ does } y \Diamond\!\!\rightarrow \sim p)))$$ [11]

Thus, it is an untouchable truth that p if and only if p and nobody is able to do anything such that, if he or she did it, it might not be the case that p. Note that van Inwagen's proposal entails that it is an untouchable truth that p if and only if p and no matter what anybody is able to do, if he or she did it, it would (still) be the case that p:

(*Can't Touch II*)

$$Up \leftrightarrow (p \;\&\; \forall x \forall y(x \text{ is able to do } y \rightarrow (x \text{ does } y \;\Box\!\!\rightarrow p)))$$

We are now, given these definitions, able to derive a variant of van Inwagen's rule alpha within our formal framework:

Rule Alpha*: $\Box p \vdash Up$

For suppose that it is necessary that p even though it is not an untouchable truth that p:

(1) $\Box p$ assumption

(2) $\sim Up$ assumption

It follows from Lewis's definition of necessity that, (even) if it were not the case that p, it would (still) be the case that p:

(3) $\sim p \;\Box\!\!\rightarrow p$ (1) and (Necessity)

It obviously follows that it is the case that p:

(4) p (3) and (Weak Centering)

For suppose that it is not the case that p. It would follow that it is and it is not the case that p (given that it would be the case that p, if it were not the case that p). I conclude that it is the case that p. However, given van Inwagen's definition of untouchability and given that it is not untouchable that p, it is either not the case that p or somebody is able to do something such that, if he or she did it, it might not be the case that p:

(5) $\sim p \vee \exists x \exists y(x \text{ is able to do } y \& (x \text{ does } y \;\Diamond\!\!\rightarrow \sim p))$ (2), Can't Touch I

It follows that somebody is able to do something such that, if he or she did it, it might not be the case that p:

(6) $\exists x \exists y(x \text{ is able to do } y \;\&\; (x \text{ does } y \;\Diamond\!\!\rightarrow \sim p))$ (4) and (5)

If we omit the particular quantifiers, we get the following result:

(7) r is able to do s & (r does $s \;\Diamond\!\!\rightarrow \sim p$) from (6)

However, there are no false counterfactuals with a necessary consequent. Hence, given that it is necessary that p, it follows that:

(8) r does $s \;\Box\!\!\rightarrow p$ (3), (Vacuity)

Hence, given Lewis's interdefinability, it follows that:

(9) $\sim (r \text{ does } s \Diamond\!\!\rightarrow \sim p)$ (8) and (Interdefinability)

Thus, the assumption that it is necessary that p even though it is not an untouchable truth that p leads to the following contradiction:

(10) $(r \text{ does } s \Diamond\!\!\rightarrow \sim p) \,\&\, \sim (r \text{ does } s \Diamond\!\!\rightarrow \sim p)$ (7) and (9)

I conclude that rule alpha* is valid within our formal framework. Thus, while van Inwagen had to *assume* that every necessary truth is untouchable, we are now able to *show* that every necessary truth is untouchable.

What is even more important, given van Inwagen's definition of being multiply able, we are now able to *show* that nobody is multiply able if all truths are untouchable (given a fairly innocuous assumption about ability). That is, given van Inwagen's definitions, we are able to argue for the following lemma:

(*Touch-Ability-Lemma*) $\forall p(p \rightarrow Up) \rightarrow \sim \exists x(x \text{ is multiply able})$

To see this, recall that somebody "is *multiply able* if there was an occasion on which he or she was trying to decide which of two (or more) incompatible alternative courses of action [...] [to perform] and was at some point in the course of those deliberations able to perform each of them" [18, p. 214]. I suppose that two courses of action, Y and Z, are only incompatible alternative courses of action if it is true that if somebody were to perform Y, he or she would *not* perform Z, and if he or she were to perform Z, then he or she would *not* perform Y. It follows that somebody is only multiply able if there are at least two courses of action, Y and Z, such that he or she is able to perform Y and able to perform Z and if he or she were to perform Y, he or she would *not* perform Z, and if he or she were to perform Z, then he or she would *not* perform Y:

(*Multiple Ability*) $\forall x\, (x \text{ is multiply able} \rightarrow \exists y \exists z\, (x \text{ is able to do } y \,\&\, x \text{ is able to do } z \,\&\, (x \text{ does } y \Box\!\!\rightarrow \sim (x \text{ does } z)) \,\&\, (x \text{ does } z \Box\!\!\rightarrow \sim (x \text{ does } y))))$

In order to show that nobody is multiply able if all truths are untouchable, I shall rely on a fairly innocuous assumption about ability, namely, the assumption that ability implies possibility:

(*Abil-Poss*) $\forall x \forall y\, (x \text{ is able to do } y \rightarrow \sim \Box \sim (x \text{ does } y))$

It is not difficult to see now that nobody is multiply able if all truths are untouchable. For suppose that all truths are untouchable and suppose, further, that somebody is multiply able:

(1) $\forall p(p \rightarrow Up)$ assumption

(2) $\exists x(x \text{ is multiply able})$ assumption

These assumptions entail a contradiction. For, given van Inwagen's definitions, these assumptions yield the following result:

(3) $\forall p(p \rightarrow \forall x \forall y(x \text{ is able to do } y \rightarrow (x \text{ does } y \,\Box\!\!\rightarrow p)))$ (1), Can't Touch II

(4) $\exists x \exists y \exists z(x \text{ is able to do } y \;\&\; x \text{ is able to do } z \;\&\; (x \text{ does } y \,\Box\!\!\rightarrow \sim (x \text{ does } z)) \;\&\; (x \text{ does } z \,\Box\!\!\rightarrow \sim (x \text{ does } y)))$ (2), Multiple Ability

Thus, if we drop the particular quantifiers, we get:

(5) $r \text{ is able to do } s \;\&\; r \text{ is able to do } t \;\&\; (r \text{ does } s \,\Box\!\!\rightarrow \sim (r \text{ does } t)) \;\&\; (r \text{ does } t \,\Box\!\!\rightarrow \sim (r \text{ does } s))$ from (4)

Suppose, however, that it is false that r does s:

(6) $\sim (r \text{ does } s)$ assumption

It follows, given that it has been assumed that all truths are untouchable, that it is an untouchable truth that it is false that r does s. Therefore, we get the following result:

(7) $\forall x \forall y(x \text{ is able to do } y \rightarrow (x \text{ does } y \,\Box\!\!\rightarrow \sim (r \text{ does } s)))$ (3) and (6)

Recall, however, that r is able to do s. It follows that:

(8) $r \text{ does } s \,\Box\!\!\rightarrow \sim (r \text{ does } s)$ (5) and (7)

Hence, it is necessarily false that r does s. For we can argue as follows:

(9) $r \text{ does } s \leftrightarrow \sim\sim (r \text{ does } s)$ (Tautology)

(10) $\sim\sim (r \text{ does } s) \,\Box\!\!\rightarrow \sim (r \text{ does } s)$ (8), (9), (Equivalence)

(11) $\Box \sim (r \text{ does } s)$ (10) and (Necessity)

However, if somebody is able to do something, then it is not necessarily false that he or she does it. Thus, a contradiction follows from the assumption that it is false that r does s:

(12) $r \text{ is able to do } s \;\&\; \sim (r \text{ is able to do } s)$ (5), (11),(Abil-Poss)

It follows that r does s:

(13) r does s (6)–(12)

However, recall that s and t are incompatible alternative courses of action. As we have already seen, this entails that if it were the case that r does s, then it would be false that r does t. It is, therefore, false that r does t:

(14) $\sim (r \text{ does } t)$ (5), (13), (Weak Centering)

However, given that it has been assumed that all truths are untouchable, it follows that it is an untouchable truth that it is false that r does t. That is, we get the following result:

(15) $\forall x \forall y(x \text{ is able to do } y \rightarrow (x \text{ does } y \;\Box\!\!\rightarrow \sim (r \text{ does } t)))$ (3) and (14)

Recall, however, that r is able to do t. It follows that:

(16) $r \text{ does } t \;\Box\!\!\rightarrow \sim (r \text{ does } t)$ (5) and (15)

It is, therefore, necessarily false that r does t. For we can, again, argue as follows:

(17) $r \text{ does } t \leftrightarrow \sim\sim (r \text{ does } t)$ (Tautology)

(18) $\sim\sim (r \text{ does } t) \;\Box\!\!\rightarrow \sim (r \text{ does } t)$ (16), (17), (Equivalence)

(19) $\Box \sim (r \text{ does } t)$ (18), (Necessity)

However, if somebody is able to do something, then it is not necessarily false that he or she does it. Thus, the assumption that somebody is multiply able even though all truths are untouchable entails a contradiction:

(20) $r \text{ is able to do } t \;\&\; \sim (r \text{ is able to do } t)$ (5), (19), (Abil-Poss)

I conclude that nobody is multiply able if all truths are untouchable:

(*Touch-Ability-Lemma*) $\forall p(p \rightarrow Up) \rightarrow \sim \exists x(x \text{ is multiply able})$

Let us now assume that the following variant of van Inwagen's rule beta is valid:[12]

Rule Beta*: $Up,\; U(p \rightarrow q) \vdash Uq$

It is now straightforward to reconstruct van Inwagen's argument within our framework. For if we let, as above, 'P_0' be a proposition that expresses the state of the world in the remote past (at a time long before the existence of the first human being), if we let 'P' be an arbitrary truth and if we let 'L' be the conjunction of all laws of nature then we can reconstruct van Inwagen's argument as follows:

(1) $\forall p(p \to \Box((p_0 \& l) \to p))$ follows from determinism

(2) p assumption

(3) $\Box((p_0 \& l) \to p))$ (1) and (2)

(4) $U((p_0 \& l) \to p))$ Rule Alpha*

(5) $U(p_0 \& l)$ premise

(6) Up (4), (5) and Rule Beta*

(7) $\forall p(p \to Up)$ (2)–(6)

(8) $\sim \exists x(x \text{ is multiply able})$ Touch-Ability-Lemma

Thus, within a quantified counterfactual logic we are able to improve upon van Inwagen's original argument given that we are able to derive important assumptions of his argument and given that we are now able to argue directly for incompatibilism.

5 *The Consequences for the Consequence Argument: Jack Spencer's Critique Revisited*

Of course, my aim is not to defend van Inwagen's argument. Instead, my aim is to develop a formal framework in which the debate about van Inwagen's argument can be conducted with rigour, clarity and transparency. What we have seen so far is that, within a quantified counterfactual logic, van Inwagen would be able to *argue* for the assumption that rule alpha is valid as well as for the assumption that his argument is an argument for incompatibilism. That is, within a quantified counterfactual logic, van Inwagen would be able to *make transparent* what one would have to reject in order to reject one of these two assumptions of van Inwagen's argument. One might raise the question, however, whether this is an important result, that is, whether this is a result that has any impact on the debate about van Inwagen's argument (other than removing doubts about van Inwagen's assumptions). In my view, this question is best approached by considering a case example: Jack Spencer's critique of van Inwagen's argument.

Jack Spencer [14] has recently argued, contrary to what is and always has been the received view, that ability does *not* imply possibility. That is, Spencer has recently argued that there might be somebody who is able to do something even though it is necessarily false that he does it. This is not the place to discuss Spencer's argument (at least in this paper, I do not want to take a stand on Spencer's argument). What is interesting, however, is that Spencer draws the conclusion that van Inwagen's argument fails because in his view, if ability does *not* imply possibility, then rule alpha is *not* valid [14, p. 483–84].

To be sure: Spencer is not very explicit about his interpretation of rule alpha [14, p. 483, fn. 25].[13] But, at least with respect to van Inwagen's latest interpretation of the argument (that is, van Inwagen's interpretation of the argument in terms of untouchability), Spencer's critique can be shown to be *misdirected* – and this becomes evident once van Inwagen's argument is reconstructed within a quantified counterfactual logic.

For what we have seen so far is that, within a quantified counterfactual logic, we have been able to derive rule alpha solely from widely held definitions and axioms of counterfactual logic *without the assumption that ability implies possibility.* Thus, at least with respect to van Inwagen's latest interpretation of the argument, Spencer is well advised *not* to reject van Inwagen's rule alpha (unless he is prepared to reject widely held definitions and axioms of counterfactual logic). If anything, Spencer is well advised to reject a different assumption of van Inwagen's argument. For what we have also seen is that, within a quantified counterfactual logic, we have been able to derive the result that, if all truths are untouchable, then nobody is multiply able. However, we have *not* been able to derive this result *without the assumption that ability implies possibility.* Thus, within a quantified counterfactual logic, it becomes evident that Spencer's critique targets the wrong premises. For the denial that ability implies possibility has *not* the consequence that there might be necessary truths that are not untouchable. If anything, it has the consequence that somebody might be multiply able even if all truths are untouchable. Thus, contrary to what Spencer predicts, the denial that ability implies possibility suggests the denial of van Inwagen's claim that his argument is an argument for incompatibilism (*not* the denial of van Inwagen's claim that rule alpha is valid).

6 Conclusion

By way of conclusion: due to the expressive limitations of his formal language, van Inwagen is unable to argue directly for the thesis that there is no free will (or that nobody is multiply able) if determinism is true. One might, therefore, want to reconstruct van Inwagen's argument within a first-order predicate logic. However, as we have seen, in order to reconstruct van Inwagen's argument within a first-order predicate logic, one would have to make use of a formal language that, given plausible assumptions about the modalities in van Inwagen's argument, leads into paradox. I suggest, therefore, to reconstruct van Inwagen's argument within a quantified counterfactual logic with propositional quantifiers. I have argued that within a quantified counterfactual logic one is not only able to argue directly for the thesis that there is no free will if determinism is true, one is also able to argue for crucial assumptions of van Inwagen's argument.

Acknowledgements

Many thanks to three anonymous reviewers of this journal for helpful comments on an earlier version of this paper.

Notes

1 Throughout this paper (at least when I do not quote other philosophers), I use capital letters (P, Q, ...) as names for propositions, I use small letters (p, q, ...) as abbreviations for sentences that express these propositions and I use small letters within quasi-quotation markers ($\ulcorner p \urcorner$, $\ulcorner q \urcorner$, ...) as names for these sentences.

2 Note that '$\Box p$' stands for 'it is a necessary truth that p'. A necessary truth, according to van Inwagen, is a truth that "*has* to be true, that would be true no matter what" [17, p. 453].

3 For suppose that, even though the world is deterministic, the conjunction of the conjunction of all laws of nature with a description of a prehuman state of the world does *not* entail all true propositions. Then, even though the world is deterministic, there is a true proposition such that it is possible that the conjunction of the conjunction of all laws of nature with a description of a prehuman state of the world is true and that proposition is false. It follows that, even though the world is deterministic, the conditional whose antecedent is the conjunction of all laws of nature and whose consequent is (the conditional whose antecedent is a description of a prehuman state of the world and whose consequent is that proposition) is *not* a necessary truth. And this is at odds with the sixth premise.

4 To see this, suppose that P is untouchable and suppose, further, that P entails Q. If P entails Q, the conditional whose antecedent is P and whose consequent is Q is a necessary truth. However, according to the first premise, every necessary truth is untouchable. It follows that P is untouchable and it follows, further, that the conditional whose antecedent is P and whose consequent is Q is untouchable.

It follows, according to the fifth premise, that Q is untouchable. Thus, the first and the fifth premise taken together entail that if P is untouchable and if P entails Q, then Q is untouchable.

5 For suppose that P is untouchable and suppose, further, that Q is untouchable. The conditional whose antecedent is P and whose consequent is (the conditional whose antecedent is Q and whose consequent is the conjunction of P and Q) is a logical truth and, therefore, a necessary truth. Thus, according to the first premise, the conditional whose antecedent is P and whose consequent is (the conditional whose antecedent is Q and whose consequent is the conjunction of P and Q) is untouchable. It follows from the fifth premise, given that P is untouchable, that the conditional whose antecedent is Q and whose consequent is the conjunction of P and Q is untouchable. Further, it follows from the fifth premise, given that Q is untouchable, that the conjunction of P and Q is untouchable. Thus, the first and the fifth premise taken together entail that if P is untouchable and if Q is untouchable, then the conjunction of P and Q is untouchable.

6 What follows is a simplified version of Montague's original argument.

7 For a similar approach with respect to paradoxes of self-reference see, for example, [5, 6, 7].

8 See, again, [5, 6, 7].

9 This axiomatic system obviously follows from Lewis's axiomatic system [8, p. 132]. For Lewis allows interchange of logical equivalents (which corresponds to Equivalence), takes any truth-functional tautology to be an axiom (which corresponds to Tautology) and adds two axioms, his fourth and his sixth axiom (which correspond to Vacuity and Weak Centering).

10 Timothy Williamson justifies this equivalence by deriving it from plausible principles about how counterfactuals relate to modality [22, p. 155–161].

11 In what follows, 'x is able to do y' is an abbreviation for 'x is or ever has been able to do y'.

12 As it turns out, this variant of rule beta can also be shown to be valid if Lewis's rule of deduction within conditionals is added to our very weak subsystem of Lewis's logic. However, given that there is an ongoing debate about the validity of this rule, I prefer to leave a discussion of this variant of rule beta for another paper. For more on this debate see, for example, [2, 11, 22].

13 At least compared to the level of sophistication that the debate about the best interpretation of van Inwagen's argument has nowadays reached. See, for example, [3, 4, 9, 12, 13, 16, 17, 18, 19]. Of course, Spencer does not want to exclude the possibility that one might come up with an interpretation of rule alpha that escapes his criticism of rule alpha [14, p. 483, fn. 25]. However, he fails to notice that such interpretations have already been proposed and, what is most important, he fails to notice that he might still criticise a different assumption of van Inwagen's argument. In my view, he fails to notice this because no formal reconstruction of van Inwagen's argument within a quantified counterfactual logic is available to him.

Marco Hausmann
Ludwig-Maximilians-Universität München (LMU Munich)
Geschwister-Scholl-Platz 1
80539 München, Germany
<marco.hausmann1@gmx.de>

References

[1] Lynne Rudder Baker. "The Irrelevance of the Consequence Argument". In: *Analysis* 68 (2008), pp. 13–22.

[2] Francesco Berto, Rohan French, Graham Priest and David Ripley. "Williamson on Counterpossibles". In: *Journal of Philosophical Logic* 47 (2018), pp. 693–713.

[3] Marco Hausmann. "The Consequence Argument Ungrounded". In: *Synthese* 195 (2018), pp. 4931–50.

[4] Michael Huemer. "Van Inwagen's Consequence Argument". In: *The Philosophical Review* 109 (2000), pp. 525–44.

[5] Saul Kripke. "Outline of a Theory of Truth". In: *Journal of Philosophy* 72 (1975), pp. 690–716.

[6] Hannes Leitgeb. "Towards a Logic of Type-Free Modality and Truth". In: *Logic Colloquium 2005: Proceedings of the Annual European Summer Meeting of the Association for Symbolic Logic, Held in Athens, Greece, July 28-August 3, 2005*. Ed. by Costas Dimitracopoulos. Cambridge University Press, 2008, pp. 28–68.

[7] Hannes Leitgeb. "What Truth Depends on". In: *Journal of Philosophical Logic* 34 (2004), pp. 155–92.

[8] David Lewis. *Counterfactuals*. Oxford: Blackwell, 1973.

[9] Thomas McKay and David Johnson. "A Reconsideration of an Argument against Compatibilism". In: *Philosophical Topics* 24 (1996), pp. 113–122.

[10] Richard Montague. “Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability”. In: *Acta Philosophica Fennica* 16 (1963), pp. 153–67.

[11] Daniel Nolan. “Impossible Worlds: A Modest Approach”. In: *Notre Dame Journal of Formal Logic* 38 (1997), pp. 535–72.

[12] Alexander Pruss. “Incompatibilism Proved”. In: *Canadian Journal of Philosophy* 43 (2013), pp. 430–37.

[13] Benjamin Schnieder. “Compatibilism and the Notion of Rendering Something False”. In: *Philosophical Studies* 117 (2004), pp. 409–28.

[14] Jack Spencer. “Able to do the Impossible”. In: *Mind* 126 (2017), pp. 465–97.

[15] Peter van Inwagen. *An Essay on Free Will*. Oxford: Clarendon Press, 1983.

[16] Peter van Inwagen. “Free Will Remains a Mystery”. In: *Philosophical Perspectives* 14 (2000), pp. 1–19. Reprinted in: Peter van Inwagen, *Thinking about Free Will*. Cambridge University Press, 2017, pp. 90–110.

[17] Peter van Inwagen. “The Consequence Argument”. In: *Metaphysics. The Big Questions 2nd ed.* Ed. by Peter van Inwagen and Dean Zimmerman, Oxford: Blackwell, 2008, pp. 450–56.

[18] Peter van Inwagen. “A Dialogue on Free Will”. In: *Methode 3* (2013), pp. 212–21.

[19] Peter van Inwagen. “Introduction: Van Inwagen on Free Will”. In: Peter van Inwagen, *Thinking about Free Will*. Cambridge University Press, 2017, pp. 1–19.

[20] Peter van Inwagen. “How to Think about the Problem of Free Will”. In: Peter van Inwagen, *Thinking about Free Will*. Cambridge University Press, 2017, pp. 149–65.

[21] Peter van Inwagen. “A Promising Argument”. In: Peter van Inwagen, *Thinking about Free Will*. Cambridge University Press, 2017, pp. 166–76.

[22] Timothy Williamson. *The Philosophy of Philosophy*. Oxford: Blackwell, 2007.