Compatibilism and the Notion of Rendering Something False

Benjamin Schnieder

Published in *Philosophical Studies* **117** (2004), 409–428.

(7.824 words, including notes & bibliography)

Abstract:

In this paper I am concerned with Peter van Inwagen's *Consequence Argument*. I focus on its probably best known version. In this form it crucially employs the notion of *rendering a proposition false*, a notion that has never been made sufficiently clear. The main aim of my paper is to shed light on this notion. The explications offered so far in the debate all are based on *modal* concepts. I argue that for sufficient results a "stronger", hyper-intensional concept is needed, namely the concept expressed by the word "because". I show that my analysis is superior to the prior ones. On the basis of this analysis I further explain why van Inwagen's argument fails.

Structure:

1. The Aim of this Paper

In this paper I will discuss van Inwagen's famous Consequence Argument and thereby the concept of someone's rendering something false.

Van Inwagen's argument which, if accepted, leads to the denial of compatibilism has attracted immense attention in the debate about free will. No wonder – it is a piece of striking conclusiveness and clarity. He presents it in his *An Essay on Free Will* in three different versions. The first of these crucially involves the notion of *rendering (a proposition) false*.[1] If someone acts in a certain way, we can describe her behaviour as rendering or making some propositions true or false. By raising my arm, I render it true, that my arm rises, and at the same time I render it false, that my arm remains motionless.

The concept is easily grasped, but it proves to be rather elusive under philosophers' scrutiny. Although it has been discussed quite a lot, no satisfying explication has yet been given – or so I will argue (section 3.d). But a proper analysis of the concept is wanting (see section 3.c-d). So I shall present such an analysis (section 4) and finally, equipped with the analysis proposed, I will block van Inwagen's strike against compatibilism (section 5). But first, in the next section, I briefly present the Consequence Argument (whoever is familiar with the argument and does not feel the need to refresh his memory may skip the following section.)

[1] The other versions of the argument, which were more prominent in recent discussions of the subject, I will not discuss (especially the so called *principle beta* stood in the limelight of philosophers' interests; see for example Fischer/ Ravazza 1996 and Crisp/Warfield 2000). Neither will I inquire into the (doubtless existing) connections between the versions. The argument's first version is of enough interest in itself, especially for its making use of the phrase "to render something false", which expresses a concept that links truth and actions.

2. The Case Against Compatibilism: The Consequence Argument

Let me briefly resume the Consequence Argument and its target. *Compatibilism* is the thesis that the truth of determinism does not imply there being no free will, where *determinism* is roughly the thesis that everything that happens follows from earlier states of the world and the laws of nature.[2] More precisely, we could say:[3]

(Df. Det) Determinism is true $\leftrightarrow_{df.}$

(i) every proposition expressing the state of the world at a time *t*, taken together with the laws of nature, implies all propositions expressing states of the world at times later than *t* &

(ii) for each time *t* there is a true proposition expressing the state of the world at *t*.

Some remarks are in order: (Df. Det) mentions states and laws of nature. Both expressions should be taken in an ordinary sense. *Laws of nature* are a certain kind of true propositions of a general form. What exactly distinguishes them from other true propositions is subject to a long debate that can't be entertained here. A *state* of some thing can be understood as the totality of all its intrinsic properties (in particular, relations to other times must be excluded).[4] So, being about to explode after five minutes does not belong to any thing's state. Otherwise determinism would be trivialised. Finally, that a proposition *expresses* a state of the world means it asserts that the world is in that state.[5] It may be that such propositions are so

[2] For the sake of brevity van Inwagen neglects forms of determinism that accept other determining sources like god or a mystic fate; I simply follow him here.

[3] See van Inwagen 1983: 65. Actually, his definition is a little stronger, because he does not restrict the implications to later times. However, (DET) suffices for his argument and I prefer the weaker version.

[4] A reflection upon the use of "state" in non-philosophical contexts suggests that essential properties may disqualify as states as well. The ordinary use of "state" seems intimately connected to ideas of *change*. Normally we do not, for example, talk about states of abstract (and unchangeable) objects like numbers, and there are some things which we call states *rather than* properties, where prototypical examples of states (hunger, drunkenness, states of aggregation) tend to be temporary characteristics of things only (cp. also Steward's illuminating remarks on states, in her 1997: especially 107-110). For the present purpose, the exclusion of certain extrinsic properties from the state of the world should, however, be sufficient.

[5] This should be specified with care. The whole point of introducing talk about states is to save determinism from collapsing into triviality. Since one could think that there are properties related to the future, we are in need of a narrower concept, that of a state. But

complex that no human being could ever come to grasp or even express them. Determinism could still be true, for it is not an epistemic but a metaphysical thesis.

The phrase "freedom of the will" has been used with several meanings. Van Inwagen is concerned with any notion of free will that satisfies the following condition:[6]

(FW) For all beings *x*:

> If *x* has freedom of the will then it is sometimes the case that *x* is able to act otherwise than she actually does.

So, the Consequence Argument is supposed to show that if determinism is true, nobody ever was able to act otherwise than she acted.

Let's turn to the argument itself. It explores the following, compelling idea:

> If determinism is true, then whatever happens follows from the laws of nature together with some truths about earlier states of the world. If at a certain time we had acted differently from the way we actually did, then either the past or the laws of nature would have had to differ from the actual. But we can change neither the past nor the laws of nature. Therefore, if determinism is true, we never had it within our powers to act in way in which we actually did not act.

This is basically the idea of the Consequence Argument. Now for the details: Van Inwagen's version of the argument involves a (very) short thought experiment: Imagine a certain judge, whom we may call Othon, who did not raise his hand at time *t*. The argument is supposed to show that under the assumption of determinism it follows that Othon was not able to raise his hand at *t*.

Let "p_o" name a proposition expressing the state of the world at a time t_0 that is earlier than Othon's birth. Let "p" be a name for a proposition expressing the state of the world at *t*.

---

notice that we can refer to one and the same state in many different ways. We can, for example, characterise a state by its relations to its preceding or succeeding states. Obviously, a proposition stating that the world is in the same state as a year ago may correctly describe the state of the world while implying a lot about the state of the world at some other time. So the notion of a proposition *expressing* a state of the world should be taken as meaning roughly: a proposition stating that the world is in such and such state, while specifying the state non-relationally.

[6] There was a time when most authors in the debate about free will would have wanted condition (FW) to be satisfied. That unanimity is gone, however, due to the much discussed problem cases produced by Harry Frankfurt.

Van Inwagen's argument can then be put as follows:[7]

(1) If determinism is true then p follows from the conjunction of p_0 and the laws of nature.

(2) It is not possible that both Othon raises his hand at t and p is true.

(3) If (2) is true then the following holds:

If Othon was able to raise his hand at t then he was able to render p false.

(4) If Othon was able to render p false and p follows from the conjunction of p_0 and the laws of nature, then Othon was able to render this conjunction false.

(5) If Othon was able to render the conjunction of p_0 and the laws of nature false, he was able to render the laws of nature false.

(6) Othon was not able to render the laws of nature false.

Therefore:

(C) If determinism is true, then Othon was not able to raise his hand at t.

Othon and his raising his hand are arbitrarily chosen and thus substitutable for whatever actor and action you like. This leads to the result that if the above argument is correct then, if determinism is true, nobody ever was able to act in a way different from how he actually acted.

The argument is definitely valid. To avoid the conclusion one thus has to deny at least one of its premises. The first two of them are uncontroversial. The remaining premises all make use of the notion of someone's rendering a proposition false. In the next section I shall make some preliminary remarks about this notion and then sketch what can be called the *standard reply* to the argument. I will argue that this reply leaves open room for manoeuvring. In sections 4 and 5 I will then try to improve upon the reply.

[7] Cf. van Inwagen 1975, 1983.

3. The Standard Reply and the Concept of Rendering Something False

a. Rendering Something false and Being Able to Render Something False

Van Inwagen introduced the notion of an agent's being able to render a proposition false as a way of describing an agent's powers, by relating her abilities to the truth-values of propositions. Similarly, but with a little more generality, we can state the following:

(Render) To say that someone rendered a certain proposition true (or false) is a way of describing one of her actions.

So, if Jean raised her arm at *t*, she rendered it true that her arm rose (at *t*) and she rendered it false, that it remained motionless at that time. The greater generality of (Render) consists in two features: (i) Van Inwagen focuses only on the power of rendering something *false*. However, truth and falsity come as two sides of one coin and whatever we say about "rendering false" should have an equivalent for "rendering true". (ii) That we can describe an action as an instance of rendering a proposition true (or false) *grounds* the possibility of describing a power for acting as the ability to render a proposition true (or false). This can be seen from the fact that the phrase "*x* is able to render *p* false" contains the relational predicate "… is able to …" as a part. This part does the job of making the phrase a description of an ability. The other part, "… render *p* false", determines the *kind* of this ability by specifying the sort of action, that the ability is an ability to undertake.

In the discussion centring around van Inwagen's argument, it is the longer phrase "*x* can render *p* false" that usually stood in the limelight. Since I will comment on some proposals made, I will in general follow this practise. However, it should be clear that the ability part of the notion might be separated, and thus an analysis of "*x* can render *p* false" implicitly provides insight into the concept expressed by "*x* renders *p* false".

b. Preliminary Clarifications: Propositions and Causality

Because a variety of things have been called by the title of a proposition, I shall briefly explain what van Inwagen takes propositions to be. Propositions in his sense are things that

(a) can be asserted, believed, doubted, etc., and that

(b) can be denoted by sentence nominalizations of the form: "that p", and that

(c) are the primary bearers of truth-values.[8]

Furthermore, they are *abstract* entities in the sense that they lack position in space and time (which means that it is senseless to predicate any temporally or spatially intrinsic properties to them), and that they do not partake in causal processes.

Now the notion of rendering a proposition false may sound as if someone could *do* something *with* a proposition, as if a proposition and its properties could be acted upon in a causal way. This idea might arise from the association with similar expressions like "making something hot", "making an omelette", or maybe "making *someone* hot". Indeed it is often the case that phrases of the form "*x* makes *y F*" indicate that *x* stands in a causal relation to *y*, as it is in the case of omelettes or of arousing someone's passion. However, to understand the phrase "*x* makes *p* true (false)", it is more instructive to think of phrases like "making someone famous". You can make someone famous without acting upon him in any way, just by acting upon other people. You can even make some dead philosopher from the nineteenth century famous, but you can surely not act upon him or exercise any causal powers on him. Rendering a proposition true (false) should be thought of like making someone famous; it is not a kind of acting upon a proposition. What you act upon are things in the world, and your acting upon them is logically sufficient for the falsehood of the proposition in question. If I, for instance, knock over a glass, I render it false that the glass keeps standing throughout the entire time of its existence. But I act only upon the glass, not upon the mentioned proposition; my acting is logically sufficient for the falsity of the proposition.

c. The Standard Reply

The Consequence Argument makes essential use of the expression "*x* can render *p* false". Thus, unsurprisingly, many attempts have been made of defining this notion. None of them, I shall argue, is really satisfying as reflecting our intuitive understanding of it. Since the notion is, regardless of its involvement in the Consequence Argument, surely of philosophical interest in its own rights, the insufficiency of the available explications already provides a reason to look for something better (a task I to which will attend in the next section). But

[8] See van Inwagen 1983: 31ff. for a characterisation of propositions.

understanding this notion is also desirable for evaluating van Inwagen's argument, as I will now show.

Though many a compatibilist felt at least irritated by van Inwagen's argument, most of them remained unconvinced. Among the many replies, the strongest ones proceed along the following lines:

> The idiom "someone is able to render the proposition that *p* false" obviously plays a crucial role in the argument. But how is this notion to be understood? Of course, the argument is only worth its money if all of its premises are true *under the same reading* of this notion. Otherwise, if there is no such reading which supports all premises at the same time, the argument fails due to an equivocation in its central term. And, unfortunately for van Inwagen, this is just how things are.

Usually, the proponents of this kind of reasoning go on to offer two different readings of the notion in question and argue that under each of them, one of the premises turns out false. Let us call this the *standard reply* to the Consequence Argument.[9]

This defence exhibits a certain weakness in resting on a *negative* existential claim which will be hard to establish in principle. However, its proponents will maintain that the burden of proof lies with their opponent. If someone claims that there is a sense of "*x* is able to render *p* true" such that the argument is sound, they will simply ask him to name this sense. That, it seems, is fair enough.

But here the defender of the argument may find some last retreat. He might insist that we in fact understand the notion and that, in a natural reading, all the premises seem to be true. Appearances may, of course, be defective. A rigorous explication of the notion thus might settle the case. But the explications offered by the authors of the standard reply, he might argue, do not catch our intuitive understanding of the notion. Thus, the case at least remains undecided.

*d. The Inadequacy of the Available Explications of "*x *can render* p *false"*

And indeed, the explications proposed in the literature are not wholly satisfying. To show this, I will take a short look at three such suggestions. In the discussion some

[9] Versions of this reply can be found, for example, in Lewis 1981 and in Fischer 1983.

desiderata for a proper analysis of the concept will arise which are not met by the proposals given.

David Lewis set forth two explications of the notion, the first of which runs as follows:[10,11]

(Lewis-1) *x* can render *p* false $\leftrightarrow_{df.}$

> *x* can do something such that, if *x* did it, an event *e* would have occurred such that □ (*e* occurs → *p* is false).

In the sense specified by (Lewis-1) we can render quite a lot of propositions false, of which it is rather surprising that we can. Most relevant for our present purpose, Lewis has argued that in this sense of "rendering false", we can after all render the laws of nature false, and that indeed no supernatural powers are needed to do so.[12] I will not repeat his argument here, since it would mainly draw away attention from my line of argument. Let us assume Lewis' argument is sound; what this would show is, among other things, that (Lewis-1) is a rather bad explication of the intuitive understanding that we have of rendering something false. For, in an intuitive sense, surely nobody (at least, no human agent) can render any law of nature false.

Well, Lewis offers another explication, which is in a way "stronger", and which indeed deprives us of the ability to render laws of nature false:

(Lewis-2) *x* can render *p* false $\leftrightarrow_{df.}$

> *x* can do something such that, if *x* did it, an event *e* would have occurred such that
>
> (i) *e* is either *x*'s action itself or an effect of this action &
>
> (ii) □ (*e* occurs → *p* is false).

---

[10] Cf. Lewis 1981: 297. He states them somewhat differently, using the notion of an event *falsifying a proposition*. However, if you apply the definition he gives of this notion, you arrive at the formulation I present.

[11] Fischer (1983: 130) proposes two similar, though not equivalent, definitions. From Lewis' definitions they differ especially in their being restricted to propositions of a certain *form*, namely propositions asserting *that a certain event is occurring*, which limits their applicability considerably. Furthermore, they do not employ a strict implication (on this point see also my footnote 15). Though I will not discuss Fischer's proposal separately, it should be clear that the criticisms I develop in what follows apply to his formulations as well.

[12] Lewis 1981: 297.

Against the intuitive correctness of (Lewis-2) van Inwagen (who subscribed to a similar definition for some time) made a good case in his *An Essay on Free Will*, which can be illustrated by the following example.[13] Let us assume that Hank Quinlan, pointing his gun at Vargas, utters "You will not survive this night". But before Quinlan pulls his trigger, he is shot by his colleague Menzies. Since no one else is interested in killing Vargas, he lives on for several years. Then the proposition that *Quinlan's last words were mistaken* (let us call it *p*) is true. However, surely someone could have killed Vargas the same night. Thereby it seems he would have rendered it true that Vargas does not survive the night, and thus he would have rendered *p* false. Quinlan would have remained right in the end. But if we follow (Lewis-2), this reasoning is mistaken. True, someone could have murdered Vargas that night; but it is not necessarily the case that if someone did so, it would have been false that Quinlan's last words were mistaken. Quinlan's choice of last words was not necessitated, at least not in any metaphysical way (even if it was so, say, in a psychological manner). So, there are possible worlds in which Vargas dies right after Quinlan, but Quinlan's last words nevertheless are mistaken; maybe because he stuttered while passing away: "8 times 7 equals 42". Van Inwagen thus adopts the following modification of (Lewis-2), in which the past is fixed by an additional clause:

(van Inwagen) *x* can render *p* false $\leftrightarrow_{df.}$

> *x* can do something such that, necessarily: if *x* does it and the past does not differ from the actual past, then *p* is false.

However, this has the surprising consequence that by leaving the room I can render it false that

(Col) Columbus has never travelled to America.

Since it is a historical fact that Columbus did travel to America, whatever I do is such that necessarily: if I do it, and the past does not differ from the actual past, (Col) is false. But this result is highly counterintuitive.[14] (Col) is false, but I do not have anything to do with this and thus I do not render it false.[15]

---

[13] The example is mine, but it works essentially in the same way as van Inwagen's example of Nostradamus, Nasser, and the Sphinx (for this see van Inwagen 1983: 67f. and Horgan 1985: 345ff. for a further discussion of the example).

[14] Van Inwagen clearly realises this drawback of his definition (see his 1983: 68). However, he thinks this unwanted feature to be a minor problem. I disagree. Van

Another curious point is that with all three definitions given so far, I can also render any necessary falsehood false.[16] I can render it false that 2=4 by scratching my neck, because necessarily, if I scratched it, it would be false that 2=4. But again it is not up to me in any intuitive sense to render it false that 2=4 or to do this with any other necessary falsehood.

From the above considerations we can extract the following desiderata as a touchstone for the adequacy of any analysis of the concept of rendering a proposition false:

(D_1) No one can render the laws of nature false.

(D_2) No one can, at *t*, render propositions expressing an aspect of the state of a world at an earlier time *t** false (whether they are true or not).[17]

Inwagen urges that a straightforward way of repairing his definition would be by adding some *ad hoc*-clauses. One could do that, of course. But such "ad hoccery" business would not illuminate our intuitive understanding of the notion. Furthermore, adding more and more clauses to the definition would make it piecemeal and hard to grasp, and it should be avoided if a simpler and more unitary explication is possible (for this, see section 4).

[15] Another available explication does without a strict implication and is based solely on a subjunctive conditional instead (see Horgan 1985: 356 fn. 16, and Fischer 1986: 256; notice that Fischer's suggestion from 1983 still differs from his proposal in 1986, since in the former, while not in the latter, he explicitly quantifies over events):

(Lewis-3) *x* can render *p* false $\leftrightarrow_{df.}$ *x* can do something such that, if *x* did it, *p* would be false.

(I call it by the name Lewis, by the way, because Horgan reports that Lewis has suggested this explication to him.) This explication faces the same problems as the above explication, (van Inwagen), by granting everybody the ability to render falsities about prior states of world past false. Furthermore, compatibilists will argue in the Lewis-style that in the sense defined in (Lewis-3) one *can* render the laws of nature false. Thus, this explication fails to capture many aspects of a natural understanding of the notion.

[16] This criticism obviously also applies to the explication mentioned in the foregoing footnote.

[17] I would prefer the simpler formulation

(D_2*) No one can, at *t*, render propositions about the world at an earlier time *t** false (whether they are true or not).

– were it not that (D_2*) is false. It is a false proposition about an earlier time that in the year 2000, no one was going to write an article about free will in the following year, but it has been made false in a time later then the year 2000. The truths about the past that cannot be rendered false (or true) must be *intrinsic* to the past; such a condition of

(D_3) No one can render necessary falsehoods false.

And, moreover, as a fourth desideratum we can hold that an adequate analysis should be capable to deal with cases like the dying Quinlan. (The same constraints hold, by the way, for the notion of rendering something true. No one can render the laws of nature or necessary truths true, nor can anyone today render propositions expressing aspects of the past states of the world true.)[18]

As I have said, a proper understanding of the notion "*x* can render *p* false" is wanting for two reasons. *First* the concept expressed is of philosophical interest in itself. *Second* the Consequence Argument owes much of its strikingness to the intuitive plausibility of its premises. An evaluation of the argument, not based on some technical notion of rendering false which departs from our intuitive understanding, is therefore desirable. As long as we lack it, the argument might be sound in its intuitive reading, however this might be spelled out.

So, in what follows I will propose an analysis, which meets the desiderata, and see what the Consequence Argument looks like in the light of it.

---

intrinsicness is integrated in van Inwagen's concepts of a *state* and of a proposition's *expressing* a state.

[18] Horgan seems to agree on the desirability of a definition meeting these desiderata, but confesses: "I myself have little idea how we might frame a single definition [with such consequences]" (Horgan 1985: 350). Well, I have got an idea about this – I will propound it in the following section.

4. A Better Explication

In the explications mentioned so far an essential aspect of the concept in question gets lost: For *x*'s ability to render a proposition *p* false one should demand that *x* can do something that is *responsible* for the falsity of *p* (responsible, of course, not in a moral sense). I suggest that we can capture this idea with the following explication:

(Df. Render) *x* can render *p* false (true) $\leftrightarrow_{\text{Df.}}$

> *x* can do something such that if *x* did it, *because of that p* would be false (true).

What is the point about (Df. Render)? Let us take a look at an example first:

If Jean raises her hand, she thereby renders the proposition that she does so true. The reason for this, I hold, is simply that it is true that she raises her hand, *because* she raises her hand.

This understanding of rendering a proposition true or false yields the wanted results that no one can render necessary falsehoods false. Nobody can do something, for example, such that *because* of his doing so it is false that 2=4. And furthermore, nobody can presently do something such that *because* of his doing so it is false that Columbus never travelled to America. Of course, *once* something could have been done about this, and Columbus in fact did it; because he travelled there, (Col) is false. Last but not least, no one can in this sense render the laws of nature of nature false. So, (Df. Render) meets our desiderata (D_1) to (D_3). It also seems to be compatible with the Quinlan case. If Vargas' life was in the hands of Mr. X that night, Mr. X could have rendered it false that Quinlan's last words were mistaken. He was able to murder Vargas, and if he'd done so, because of that Quinlan's last words would have been true. The counterfactual used in (Df. Render) fixes the past in the necessary way.

What is most important about (Df. Render) is the use of the connective "because". One might wonder how and even whether its meaning could be spelled out satisfactorily.

Now, first and foremost, even if we could not do this, that would not count against my analysis. What is important is that we are competent users of the connective "because", competent enough so that my analysis yields the desired

results. If we cannot explain our usage this does not force us to deny its meaningfulness.

And admittedly, I cannot offer a cut-and-dried definition of this connective. Indeed, I tend to regard it as primitive.[19] But taking it as undefinable does not mean that we cannot say anything illuminating about it. And fortunately, there is quite a lot to say about the meaning of "because". Generally speaking, "because" signifies an *explanatory relation*. Such a relation may be causal, as in the case of the following sentence:[20]

(J) Because Jean was in a muddled state of mind she grassed on Belmondo to the police.

But causal explanation is just one kind of explanation. There are many cases where it is possible to use "because" without implying causality, as in:

(E_1) He should be punished because he committed a crime.

(E_2) Thorsten is my brother-in-law because he's married to my sister.

(E_3) There are no round squares because the concept of a round square involves a contradiction.

(E_4) 5 is prime, because neither 2, 3, nor 4 are divisors of it.

In commenting on (Df. Render), I made the following claim: It is true that Jean raises her hand, because she raises her hand. Hereby I employed a fundamental idea about the concept of truth that goes back to Aristotle who wrote:

---

[19] Especially, I hold that "because" cannot be defined in purely modal terms. Neither strict implications nor subjunctive conditionals imply correlated statements because-statements, as a short reflection upon some mathematical examples will show. And furthermore, a statement "*p* because *q*" does not have straightforward modal implications. It does not imply a strict conditional, since it allows for *contingent* explanations (which is the usual case for causal statements). And while it may *support* the subjunctive "if it were not the case that *q*, it would not be the case that *p*", it does not *imply* it (as well-known cases of over-determination show). Thus, concentrating solely on modal notions for an explication of "*x* can render *p* false" can be seen as the *proton pseudos* of the accounts discussed.

[20] I hereby do not intend to settle the question about what the proper relata of the causal relation are. They might be facts, things that are designated by whole sentences (cp. Mellor 1987, Bennett 1988: Ch. IX). They might as well be events. But even if they are, the "because" in (J) would signify *a* causal relation, namely the relation of causal explanation (cp. Davidson 1967).

> It is not because we think that you are white, that you *are* white, but because you are white we who say this have the truth.[21]

Abstracting a little from his words, Aristotle's insight can be generalised with the following theorem:[22]

(T) $\forall p$: If it is true that p at all, then it is true that p, because p.[23]

Theorem (T) bears some similarity to Tarski's T-scheme which, in its homophonous formulation, reads as follows:

(TS) "s" is true iff s.

But of course there are many important differences between (T) and (TS). I will mention only the difference which is most important in our present context: (TS) makes use of the sentential connector "iff", a connector that indicates a *symmetrical* relation. Not so "because". Indeed, from the truth of "p because q" it follows the falsity of "q because p".[24]

Back to (T): The explanatory role that "because" plays in (T) is a conceptual one, not a causal. Such a conceptual explanation might consist in showing a

---

[21] Aristotle, *Metaphysics*, 1051b6-8. The translation follows Barnes 1991: 1661. A note in passing: Van Inwagen himself alludes to this quotation when he comments on the notion of truth (van Inwagen 1983: 33f.).

[22] Bernard Bolzano developed an account of grounding, a relation signified by the connective "because", in which this principle plays an important role (see his *Wissenschaftslehre*, especially §198, and furthermore Tatzel, forthcoming, for a careful reconstruction of Bolzano's views).

[23] As you might have noticed, the quantifiers in this formula should be treated with care. The variables do not stand in the position of a singular term but rather in *sentence* position. Under the common, objectual reading of quantifiers, the formula would thus collapse into ungrammatical non-sense. To avoid this, we could give them a *substitutional* reading. For my present purpose I may leave it like that. But I should note that I prefer a rather non-standard alternative. Read substitutionally, the formula will lose an essential part of its generality, for substitutional quantification is essentially dependent upon the availability of certain linguistic forms. The alternative would be to accept a *third* kind of quantification, quantification into sentence (or general term) position which is not substitutional. I cannot defend such an account here in detail (but see Künne 1992: 232f., Williamson 1999: 259ff., and Simons 1997: 263 for some further remarks on this point).

[24] Here I presuppose that explanation is *a*-symmetric and not merely *anti*-symmetric, i.e. I exclude the possibility of self-explaining truths. Even if explanation were merely antisymmetric, it would crucially differ in this respect from the "iff" and even from its modally strengthened version "□ ... iff ...", since indisputably these connectors are *not* antisymmetric.

conceptual entailment (which is, I guess, the case in (E_3)). But sometimes it consists in something more. Take a look at (E_2). What it gives us, it seems, is a *partial explanation* of the concept of a brother-in-law. Giving an explication of a concept is more than showing a conceptual entailment. Herein lies the important core of Aristotle's insight. With (T) we are given a partial explanation of the concept of truth itself. The necessary equivalence of the proposition that *p* and the proposition that it is true that *p* is essential to our understanding of the concept of truth.

But is there not a certain air of pointlessness about (T)? If someone asks

(Q) Why is it true that snow is white?

and he is offered the answer:

(A) Well, because snow is white.

he will probably be disappointed, and rightly so. He might be reminded of Shakespeare's Lucetta who confessed, "I have no other, but a woman's reason; I think him so, because I think him so".[25] I may add that, not only for the sake of political correctness, it is more appropriate to speak of a *fool*'s reason here.

Now, this may be conceded without having to give up (T) in consequence. Since the necessary equivalence of "*p*" and "it is true that *p*" is *evident* (for anybody understanding the sentences), we often treat them as interchangeable. Someone who asks "Why is it true that snow is white?" may be happy to reformulate his question just as

(Q-2) Why is snow white?

To this latter question, of course, "Because snow is white" is not an appropriate answer. But although the equivalence "*p*" and "it is true that *p*" might be evident, we still have to recognise two different propositions expressed here (only one of them involves the concept of truth). And so, one might ask question (Q), intending to ask *exactly this* question *rather than* (Q-2). One might do so for example by stressing the word "true". Then the correct answer will be (A), however boring it may be.

[25] *The Two Gentlemen of Verona*, Act 1, Scene 2.

Still, one might be unsatisfied. If something is explained, one might say, some alternative possibilities must be excluded.[26] To this I would respond that this is often the case, but simply not always. If you are told that Thorsten is my brother-in-law because he's married to my sister, as long as you forget about the possibility of homosexual marriages, no possibility is excluded for you. Nevertheless, (E_2) seems to be a perfectly true, *explanatory* statement. Also in (E_4) I cannot see what other possibilities could be there for 5 to be a prime number, except the one indicated.[27]

Let us look again at (Df. Render) now. What kind of explanatory relation is relevant for it? We have to distinguish several cases: What is always involved is the conceptual relation expressed in (T). The most basic instances of (Df. Render) are cases in which a proposition, namely the proposition that *x* φ-s, is rendered true by *x*'s φ-ing itself:

(R) An agent *x* renders the proposition that *x* φ-s true, if *x* φ-s. She does so for the reason that it is true that *x* φ-s, because *x* φ-s.

But there are more complicated instances, in which apart from the conceptual a causal relation is involved. If Othon raises his arm and by doing this he hits a glass off the table, then he renders it true that the glass falls off the table. Why? Othon's raising his arm is the cause of the falling of the glass, so: because Othon raises his arm, the glass falls from the table. And because the glass falls, it is true that it does. Since, by and large, "because" is a transitive connector, Othon does something such that because of his doing so, it is true that the glass falls. Therefore he renders it true that it does. Not surprising, indeed, but fine. (Df. Render) delivers the goods. Let me, before ending this paragraph, mention that of course besides the explanatory force of (T) and some causal ones also other conceptual explanations might play a role for instances of (Df. Render). A woman gives birth to a child, and by doing it she renders it true that her mother becomes a granny. Why? It is true that her mother becomes a granny, because she becomes one. And she becomes one,

---

[26] The indicated worry may arise from adherence to the *Contrast Theory* of Why-Questions (and thus, of explanations), of which van Fraassen is a prominent defender (see, for example, van Fraassen 1980: especially 127).

[27] See furthermore Dennis Temple (1988), who has argued that the difference between the Contrast Theory of Why-Questions, which holds that the proper form of a why-question is "Why *p* (rather than *q*)?", and the somewhat classic, propositional account, which takes the form "Why *p*?" as the default form of why-questions, is not that big after all. For any why-question containing contrasted alternatives ("*p* rather than *q*") may be seen as asking for an explanation of the proposition expressed by "*p* and not *q*".

because the woman gives birth to the child. The latter use of a "because" expresses a conceptual explanation focussing on the concept of a granny.

5. The Case Against the Consequence Argument

Now let us go back to van Inwagen's argument and look at the premises again, keeping in mind the given explication of rendering false. Usually, advocates of the standard reply want to deny one of the premises (5) and (6).[28] But on the account proposed, these statements seem to be true after all. Nothing could be done such that because of it the laws of nature would be false. Furthermore, nobody can do something such that because of that a proposition expressing a former state of the world would be false. But then it is arguable that, if anyone can do something, because of which a conjunction of such a proposition and another proposition p would be false, then indeed because of his doing it, p would have be false.[29] Thus my analysis explains van Inwagen's intuitions about the truth of these premises quite well.

So far so good for the argument; it fails nevertheless, because of premise (4):[30]

(4) If Othon was able to render p false and p follows from the conjunction of p_0 and the laws of nature, then Othon was able to render this conjunction false.

(4) follows from the general principle:

(4-P) For all beings x and all propositions p, q:

If x is able to render p false and p follows from q, then x is able to render q false.

And (4-P) looks plausible at first; indeed, van Inwagen writes about it:

> This principle seems to be analytic. For if Q entails R, then the denial of R entails the denial of Q. Thus, any condition sufficient for the falsity of R is also

[28] See, for example, the already mentioned Lewis 1981, Fischer 1983, Horgan 1985, and Fischer 1986.

[29] This is the general principle on which van Inwagen relies in defending (5); cp. van Inwagen 1983: 72f.

[30] Notice that here I depart from the way of the standard reply, since typical proponents of it take premise (4) for granted.

> sufficient for the falsity of *Q*. Therefore, if there is some condition that *S* can produce that is sufficient for the falsity of *R*, there is some condition (that same condition) that *S* can produce that is sufficient for the falsity of *Q*. (van Inwagen 1975: 192)[31]

I agree with van Inwagen that (4-P) *seems* to be analytic. However, it is not; it is simply false. Something is wrong with his argument. If an agent produces a condition that is sufficient for the falsity of *p*, and *p* follows from *q*, then she produces a condition that is *sufficient* for the falsity of *q*. But it does not follow that *q* is false *because* of this condition. And therefore it does not follow that the agent thereby renders *q* false; sufficiency is simply not enough.

My argument against (4-P) is in a nutshell that the sentential connective "because", which is involved in the correct explication of the notion of rendering something false, is not closed under logical implication. And therefore, "rendering false" itself is not closed under logical implication. But that it is, is presupposed in (4-P), and this is the reason why (4-P) is false.

The following examples will illustrate this claim (and provide counter-examples to (4-P)):

(i) According to the classic notion of implication, a necessary falsehood implies any proposition whatsoever. Let *p* be a proposition that Jean can render false. *P* is implied, for instance, by the proposition that 2=4. If "because" were closed under implication, from the fact that Jean can render *p* false, it would follow that Jean can render false the proposition that 2=4. But as we said before, nobody can ever render this proposition false.

(ii) Even if there are some free actions there still could be some determined actions. So let us suppose that a certain action, Jean's kissing Belmondo for example, was determined. Nevertheless Jean rendered it true that she kissed Belmondo, because she was doing so. Because she rendered it true, she also was able to do so.

Now we can imagine a pair of propositions implying that Jean does not kiss Belmondo. One of them might be a false description of the remote past, p_f, the other a false description of the laws of nature, l_f. Therefore it holds:

(1) Jean can render false the proposition that she does not kiss Belmondo.

---

[31] Cf. van Inwagen 1977: 94 f. and van Inwagen 1983: 72.

(2) That Jean does not kiss Belmondo is implied by the conjunction of p_f and l_f.

If (4-P) were true, it would follow:

(3) Jean can render the conjunction of p_f and l_f false.

But this is false. Nothing that Jean can do is such that because of her doing so, the conjunction of p_f and l_f would be false. This proposition is false for different reasons, and Jean can do nothing about it.

Let me take stock: We have seen that, in its natural reading, van Inwagen's argument rests on a false principle. It is false because it (implicitly) presupposes the logical closure of "because". That it does could not be seen before an adequate analysis of the notion of rendering something false was found. By presenting such an analysis I not only showed *that* van Inwagen's argument fails, but also explained *why* it does so.

With all I said, the question about the truth or falsehood of compatibilism remains untouched. But one thing is clear: Compatibilism is nothing that anyone could render false. Not because it is true (and, by the way, I do not know if it is true). But whether it is or not – its truth-value is at any event beyond our powers to affect.[*]

[*] I am deeply indebted to Wolfgang Künne, whose lectures on truth were an important source of inspiration for my views in this paper. An earlier version of the paper has been presented at the *Mind and Action III* conference (Lisbon 2001) and at the little *Ockham-Society* in Oxford; in both cases I'm thankful for the positive reactions and comments from the audience. Last but not least, I received helpful comments from John Hyman, Roger Teichmann, and an anonymous referee of this journal, and I wish to express my gratitude for that.

References

Aristotle: *The Complete Works of Aristotle*. Edited by J. Barnes, Priceton: Princeton University Press, 4th edition, 1991.

Bennett, Jonathan (1988): *Events and Their Names*, Indianapolis: Hackett.

Bolzano, Bernard: *Wissenschaftslehre* (in four volumes); Reprint, Leipzig 1981.

Crisp, Thomas & Warfield, Ted (2000): "The Irrelevance of Indeterministic Counterexamples to Principle Beta", *Philosophy and Phenomenological Research* 61, 173–84.

Davidson, Donald (1967): "Causal Relations", in: Davidson, D. (1980), *Essays on Actions and Events*, Oxford: Clarendon Press, 149–62.

Fischer, John Martin (1983): "Incompatibilism", *Philosophical Studies* 43, 127–37.

Fischer, John Martin (1986): "Van Inwagen on Free Will", *Philosophical Quarterly* 36, 252–60.

Fischer, John Martin & Ravizza, Mark (1996): "Free Will and the Modal Principle", *Philosophical Studies* 83, 213–30.

Horgan, Terence (1985): "Compatibilism and the Consequence Argument", *Philosophical Studies* 47, 339–56.

Künne, Wolfgang (1992): "Bolzanos blühender Baum – Plädoyer für eine nicht-epistemische Wahrheitsauffassung", in: Forum für Philosophie Bad Homburg (ed.): *Realismus und Antirealismus*, Frankfurt am Main: Suhrkamp, 224–44.

Lewis, David (1981): "Are we Free to Break the Laws", in: Lewis, David (1986): *Philosophical Papers* II, Oxford: Oxford University Press.

Mellor, D. H. (1987): "The singularly affecting facts of causation", in: Mellor, D. H. (1991): *Matters of Metaphysics*, Cambridge: Cambridge University Press, 201–24.

Simons, Peter (1997): "Higher-Order Quantification and Ontological Commitment", *Dialectica* 51, 255–71.

Steward, Helen (1997): *The Ontology of Mind*, Oxford: Clarendon Press.

Tatzel, Armin (2002): "Bolzano's Theory of Ground and Consequence", *Notre Dame Journal of Formal Logic* 43, 1–25.

Temple, Dennis (1988): "The Contrast Theory of Why-Questions", *Philosophy of Science* 55, 141–51.

Van Fraassen, Bas (1980): *The Scientific Image*, Oxford: Clarendon Press.

Van Inwagen, Peter (1975): "The Incompatibility of Free Will and Determinism", *Philosophical Studies* 27, 185–99.

Van Inwagen, Peter (1977): "Reply to Narveson", *Philosophical Studies* 32, 89–98.

Van Inwagen, Peter (1983): *An Essay on Free Will*, Oxford: Clarendon Press.

Williamson, Timothy (1999): "Truthmakers and the Converse Barcan Formula", *Dialectica* 53, 253–70.