Mar 28th, 2:00 PM - 2:30 PM

# Van Inwagen's modal argument for incompatibilism

Katerina Psaroudaki
*University of Nebraska-Lincoln*

# VAN INWAGEN'S MODAL ARGUMENT FOR INCOMPATIBILISM
Katerina Psaroudaki

## INTRODUCTION

Incompatibilism is the metaphysical view according to which, determinism is incompatible with free will; if determinism is true then humans don't have free will or, alternatively, if humans possess freedom of will then determinism is false. On the other side of the debate, compatibilists argue that freedom of will is compatible with determinism; we could exercise our free will even if the future was totally determined by the past. Van Inwagen, in his paper "A Modal Argument for Incompatibilism", puts forward a very compelling argument against compatibilism according to which, if we don't "have a choice" about whether determinism is true nor do we "have a choice" about whether the past and the laws of nature are true then necessarily we don't "have a choice" about the predetermined future either. In the present paper, I shall try to show that his argument doesn't work by attacking the first premise. In arguing for that, I will proceed as follows: in section (I) I will briefly present the modal argument, in section (II) I will explain why the first premise is false or how the compatibilist could refute it under one possible understanding of it, and in section (III) I will attempt a different interpretation, which also fails to provide well-justified support to the first premise. One thing to keep in mind is that I won't argue in favor of compatibilism and some of the ideas developed in this paper may count against it. My goal is modest: to demonstrate that Van Inwagen's modal argument in favor of incompatibilism is flawed under the two interpretations of its first premise provided here.

## (I) VAN INWAGEN'S ARGUMENT

Van Inwagen understands the Consequence Argument as follows:

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 (2013, 117).

The modal formulation of the argument is the following:

(1) $\Box ((P_0 \ \& \ L) \rightarrow P)$.

(2) $\Box (P_0 \rightarrow (L \rightarrow P))$.

(3) N ($P_0 \rightarrow$ (L→P)).

(4) $NP_0$.

(5) NL.

(6) Therefore, NP.

Let us now explain what each of the variables in the modal argument denotes. '$P_0$' is an abbreviation of a proposition representing the state of the world at a particular instant in the remote past. 'L' is an abbreviation of a proposition representing the conjunction of all the laws of nature. 'P' is an abbreviation of a proposition describing the state of the world at a particular instant in the future. The operator '□' represents broadly logical or metaphysical necessity and, lastly, the operator 'N' is an abbreviation of the proposition 'No one has, or ever had, any choice about whether'.

Let us next proceed with the justification of each premise. The first premise is, according to Van Inwagen, a logical consequence of determinism; it follows necessarily from the truth of determinism. The second premise follows from the first by the application of elementary modal and sentential logic. The third premise follows from the second by the application of rule *alpha* '□p →Np': 'No one has, or ever had, a choice about what is necessarily true'. The fourth premise is true since no one has any choice about what happened in the remote past: "The proposition that $P_0$ is a proposition about the remote past. We could, if we like, stipulate that it is a proposition about the distribution and momenta of atoms and other particles in the inchoate, presiderial nebulae. Therefore, surely, no one has any choice about whether $P_0$" (120). Likewise, premise (5) is true since no one has, or ever had any choice, about whether the laws of nature are true. The laws of nature, according to Van Inwagen, are basically laws of physics: "If it is a law of nature that angular momentum is conserved, then no one has any choice about whether angular momentum is conserved" (120).

The conclusion follows from premises (3), (4) and (5) by the application of rule *beta* 'N(p→q), Np → Nq': N($P_0$→(L→P)) entails $NP_0$ →N(L→P) which entails N(L→P) given that $NP_0$ is true, which further entails NP given NL and a second application of rule *beta*. Van Inwagen argues that the rule of inference beta is valid since he cannot think of any counterexamples where someone does not have a choice about whether 'p→q' nor does he have a choice about whether 'p', but, paradoxically, he *does* have a choice about whether 'q'. For

example, if one has no choice about whether 'If the sun explodes in 2000 AD, then all life on earth will end in 2000 AD, and he, also, has no choice about whether 'The sun will explode in 2000 AD', then it seems to be logically impossible that he *does* have a choice about whether 'All life on earth will end in 2000 AD'. Nevertheless, if determinism is compatible with freedom of will, Van Inwagen concedes, it's possible that one *has* a choice about whether a proposition representing a future state of affairs in the world is true even though he has no choice about whether a proposition describing the remote past is true nor does he have a choice about whether the logical implications of determinism are true. But, he wants to test the validity of rule *beta* against counter-instances that do not presuppose the compatibility of freedom of will and determinism (123); as long as no such instances can be found, rule *beta* is valid.

To sum up, if we don't have a choice about whether 'The remote past and the conjunction of the laws of nature entail the future', nor do we have a choice about the remote past and the conjunction of the laws of nature, then we don't have any choice about the future either. If freedom of will is defined as 'the ability to do otherwise' and determinism as the metaphysical view that 'the past and the laws of nature entail the future', Van Inwagen has shown that, determinism is incompatible with free will since it it rules out our ability to change the future and act differently than expected and predetermined.


## (II) FIRST INTERPRETATION OF THE PREMISE:
### *NECESSARILY,* THE PAST AND THE CONJUNCTION OF THE LAWS OF NATURE ENTAIL THE FUTURE

In the present section I will a) suggest a possible interpretation of the first premise of the modal argument and b) explain why the premise is false provided that the suggested interpretation is correct. To remind you, Van Inwagen claims that if determinism is true, then it follows that '$\Box ((P_0 \& L) \rightarrow P)$'. One way to understand that is by identifying determinism with what is inside the parenthesis and explain how broadly logical necessity (the box outside the parenthesis) follows from the truth of determinism. In other words, we can assume that determinism is the metaphysical view that 'the past and the conjunction of the laws of nature entail the future' $((P_0 \& L) \rightarrow P)$ and then argue for the inference: if determinism is true, then '*Necessarily,* the past and the conjunction of the laws of nature entail the future' ($\Box ((P_0 \& L) \rightarrow$

P) which is equivalent to saying that it is true in all metaphysically possible wolds that the past and the laws of nature entail the future. In what follows, I will provide three distinct arguments that could support the inference from the truth of determinism to the truth of the first premise (construed as such) and show how each one fails to do so.

(i) *The fatalist fallacy*

The first possible argument underlying the inference from the truth of determinism to the truth of the first premise could be reconstructed as follows:

(1) $\square$ (If determinism is true, then $(P_0 \ \& \ L) \rightarrow P$).

(2) Determinism is true.

(3) Therefore, $\square$ $((P_0 \ \& \ L) \rightarrow P)$.

The first premise is true by definition and given the present interpretation that identifies determinism with what is inside the parenthesis ($D = (P_0 \ \& \ L) \rightarrow P$): it's necessarily true (or true in all metaphysically possible worlds) that if determinism is true then its conceptual and logical implications are also true. The second premise is true by assumption. Does the conclusion, though, logically follow from the premises? Certainly not; the above argument is logically fallacious. To see why, consider the following similarly structured argument:

(1) $\square$ (If John is a bachelor, then John is an unmarried man).

(2) John is a bachelor.

(3) Therefore, $\square$ (John is an unmarried man).

It can't be true that 'Necessarily, John is an unmarried man' on the assumption that he is (contingently) a bachelor, despite the broadly logical necessity of the conceptual truth 'If John is a bachelor, then John is an unmarried man'. On a similar vein, it can't be true that '*Necessarily*, the past and the conjunction of the laws of nature entail the future' on the assumption that determinism is true, even though it's necessarily true that 'If determinism is true, then the past and the conjunction of the laws of nature entail the future'. The argument is fallacious and thus cannot justify the inference from the assumption that determinism is true to the conclusion that '$\square$ $(P_0 \ \& \ L) \rightarrow P)$'.

(ii) *What if determinism was necessarily true?*

The above argument could be reconstructed in a logically valid form by replacing the second premise with the assumption the determinism is *necessarily* true:

4

(1) □ (If determinism is true, then ($P_0$ & L) → P).

(2) □ (Determinism is true).

(3) Therefore, □ (($P_0$ & L) → P).

Even though the argument is logically valid, we have good reasons to reject it by attacking the second premise. For one thing, Van Inwagen never makes the claim that determinism is necessarily true. For another, being a compatibilist does not commit one to the metaphysical necessity of determinism; Vihvelin, for instance, argues that determinism is an empirically contingent proposition that may be true in some metaphysically possible worlds but not in others, and that the debate between compatibilists and incompatibilists was initially fueled by the empirical success of science in making accurate predictions about the future (2013, 26). In addition, the debate between determinists and indeterminists seems to be inextricably tied with the recent discoveries in theoretical physics, with some determinists supporting their metaphysical views by appeal to deterministic natural theories (such as Newtonian Mechanics and Chaos theory) while indeterminists justifying their views on the basis of indeterministic theories in Quantum Mechanics (Hoefer, 2010).

Granted all that, the burden of proof lies with the incompatibilist, who would like to pursue the present line of reasoning, to show how determinism could be understood as a metaphysically necessary proposition (such as the proposition 'no one is taller than himself' or 'if something is red, then it is also colored'), and, by so doing, convince the compatibilist who thinks that determinism is an empirically contingent view. Until then, the argument at issue, despite being logically valid, fails to establish its conclusion that '*Necessarily*, the past and the conjunction of the laws of nature entail the future'.

*(iii) 'Determinism is true' and 'The past and the conjunction of the laws of nature entail the future' are logically equivalent propositions*

Consider the following argument in support of the proposition '□ (($P_0$ & L) → P)':

(1) If determinism is true, then □ (($P_0$ & L) → P).

(2) Determinism is true.

(3) Therefore, □ (($P_0$ & L) → P).

The argument is logically valid. To see what's wrong with it, consider the following argument that exhibits the same logical structure:

(1) If John is a bachelor, then □ (John is an unmarried man).

(2) John is a bachelor.

(3) Therefore, □ (John is an unmarried man).

The argument is equally logically valid but it's not sound because the first premise is false: even though in natural language the proposition 'If John is a bachelor, then necessarily John is an unmarried man' makes absolute sense, in Modal logic the correct formulation of the proposition is '□ (If John is a bachelor, then John is an unmarried man)'. Similarly, we can justifiably utter the proposition 'If determinism is true, then necessarily the past and the conjunction of the laws of nature entail the future', but the modal equivalent of the proposition should be '□ (If determinism is true, then the past and the conjunction of the laws of nature entail the future'. In that case, the argument would collapse to the fatalist fallacy discussed and dismissed in the first subsection.

To put the objection to the present reconstruction of the argument in a different perspective, we could say that logically equivalent propositions are either both contingent or both necessarily true. The propositions 'John is a bachelor' and 'John is an unmarried man' are logically equivalent since they imply each other: if John is a bachelor, then John is an unmarried man and if John is an unmarried man, then John is a bachelor. As a result, if the proposition 'John is a bachelor' is contingently true, so is the proposition 'John is an unmarried man'. Likewise, the propositions 'Determinism is true' and 'The past and the conjunction of the laws of nature entail the future' are logically equivalent: if determinism is true, then the past and the conjunction of the laws of nature entail the future and vice versa. Therefore, if determinism is contingently true, so is its logically equivalent proposition '$(P_0 \& L) \rightarrow P$'. The '□' operator in front of the latter proposition would be justified iff determinism was necessarily true, and, as earlier shown, that is an assumption that not every compatibilist would endorse.

To recapitulate the objections presented so far in this section: determinism could be understood either as contingently or as necessarily true. If Van Inwagen understood determinism to be an empirically contingent view, then he could not show that its logically equivalent proposition 'The past and the conjunction of the laws of nature entail the future' is necessarily true without committing a logical fallacy. If, on the other hand, he understood determinism to be

necessarily true, he would have a hard time convincing the compatibilists, who consider determinism to be empirically contingent, to accept an argument against their view based on an assumption that they clearly reject. Either way, the proposition '$\Box ((P_0 \text{ \& } L) \rightarrow P)$' does not justifiably follow from the proposition that determinism is true.

## (III) SECOND INTERPRETATION OF THE PREMISE:
## THE PAST AND THE CONJUNCTION OF THE LAWS OF NATURE *LOGICALLY* ENTAIL
## THE FUTURE

In this section I shall try to provide a different interpretation of the first premise of Van Inwagen's modal argument runs as follows: the '$\Box$' operator, does not denote *broadly logical or metaphysical necessity* in the sense that the past and the conjunction of the laws of nature entail the future in all metaphysically possible worlds; rather, it denotes *logical entailment* in the sense that the past and the conjunction of the laws of nature *logically* entail the future exactly as the premises of a deductively valid argument *logically* entail its conclusion. For illustration, consider the following argument:

(1) If Socrates is human, then Socrates is mortal.

(2) Socrates is human.

(3) Therefore, Socrates is mortal.

The argument is logically valid for if the premises are true then the conclusion is necessarily true. In a sense, the conjunction of the two premises *logically* entails the conclusion or, to put it in a modal form, '$\Box (P(1) \text{ \& } P(2) \rightarrow C(3))$'. In other words, it's necessarily true that the conjunction of the propositions 'If Socrates is human, then Socrates is mortal' and 'Socrates is human' entails the proposition 'Socrates is mortal'. Likewise, if determinism is true (even contingently), then it's necessarily true that any proposition describing the world at a particular instant in the remote past along with a proposition representing the laws of nature in their totality *logically* entail a proposition describing the world at any particular instant in the future. Unlike the first interpretation which identifies determinism with the entailment inside the parenthesis ($D= (P_0 \text{ \& } L) \rightarrow P$), the current interpretation considers the whole premise, including the box operator, to be the definition of determinism ($D= \Box ((P_0 \text{ \& } L) \rightarrow P)$).

To make this interpretation work, we need to reconstruct determinism as a deductively

7

valid argument. As a first approximation, and given what Van Inwagen says about the relevant propositions, the argument could be reconstructed along the following lines:

(1) $L_1$ & $L_2$ & $L_3$ & ... & $L_v$.

(2) $P_0$ = A description of the distribution and momenta of atoms and other particles in the inchoate presiderial nebulae.

(3) Therefore, $P$ = Jones will vote for the Democrats in the upcoming elections & $P_1$ & $P_2$ &… & $P_v$.

The first premise unifies the laws of nature into a single proposition. The laws of nature are basically the laws of physics or natural sciences in general. In his paper 'The Incompatibility of Free Will and Determinism', Van Inwagen uses the term 'laws of physics' instead of 'laws of nature' since psychological voluntaristic laws fall under the latter and his argument applies only to the former: "I should not want to use, instead of 'laws of physics', some term like 'laws of nature' that might legitimately be applied to voluntaristic laws. Thus, for all that is said in this paper, it may be that some version of determinism based on voluntaristic laws is compatible with free will" (1974, 187). In the paper under discussion he uses the term 'laws of nature' but given the single example that he gives about the law of angular momentum conservation and the justification for premise (5) ('NL') which becomes almost uncontroversial if the laws 'we have no choice about' are laws of physics, we could safely draw the conclusion that the first premise of the *deterministic* argument is equivalent to a proposition representing the conjunction of all the 'laws of physics'. As far as the other two premises are concerned, the second is a description of the world at a particular instant in the remote past and the conclusion is a description of the world at a particular instant in the future consisting of a conjunction of different propositions ($P$, $P_1$, $P_2$,…, $P_v$) representing distinct but coexisting states of affairs in the world.

One condition that a deductively valid argument needs to meet is the following: the terms appearing in the conclusion need to be equivalent in meaning with the terms appearing in the premises. For example, the terms 'Socrates' and 'mortal' preserve their meaning throughout the argument given earlier. If 'Socrates' referred to the ancient Greek philosopher in the first premise and, say, a horse in the conclusion, the argument wouldn't be logically valid. The same logic must apply to the *deterministic* argument: we need to make sure that each of the terms appearing in the description of the future world has an equivalent term in either the first premise

representing the conjunction of all the natural laws or the second premise representing the physical state of the world at a particular instant in the remote past. Some propositions about the future, though, such as the proposition 'Jones will vote for the Democrats in the upcoming elections', include terms ('vote', 'Democrats', 'elections') that do not figure in the laws of nature nor in the physical description of the world in the remote past. Therefore, the past and the conjunction of the natural laws do not *logically* entail all propositions about the future since some terms appearing in the latter do not have a logical equivalent in the former, and, thus, the present interpretation fails to establish a logically plausible first premise. In what follows, I will consider three objections to my argument.

(i) *Why not reduce the natural kind predicates in the conclusion to the natural kind predicates in the first premise?*

Someone might argue that one way to make the argument deductively valid is by reducing some of the terms figuring in the future description of the world to the natural kind predicates figuring in the laws of nature as represented in the first premise. The problem, though, with the present suggestion is that for some propositions, such as 'Jones will vote for the Democrats in the upcoming elections', no corresponding physical description could be given that would logically (semantically) subsume the proposition under some natural kind predicate appearing in any of the natural laws depicted in the first premise. What explains the irreducibility of natural kind predicates of the social sciences ('vote', 'elections', etc.) to natural kind predicates of the physical sciences is, following Jerry Fodor, the multiple physical realization of social events.

In his paper 'Special Sciences (or: The Disunity of Science as a Working Hypothesis)' (1994), Jerry Fodor specifies one condition that is necessary for a successful reduction of the laws of special sciences (psychology, economics, anthropology e.t.c.) to the laws of physics: the formulation of bridge laws that connect each predicate of the special sciences laws with a predicate of the physical laws. Suppose that '$S_1x \rightarrow S_2x$' is a law of the special sciences and '$P_1x \rightarrow P_2x$' is a law of physics, then in order to reduce the former to the latter we need to establish two bi-directional bridge laws: '$S_1x \leftrightarrow P_1x$' and '$S_2x \leftrightarrow P_2x$'. Fodor argues that such bridge laws connecting special and natural predicates are hard to establish due to the multiple physical realizability of the events investigated by the special sciences. For example, the

9

predicate 'monetary exchange' appearing in Gresham's law cannot be reduced to a natural predicate since the event described by the special predicate could be realized by many distinct physical events involving dollar bills, strings of wampum, signing one's name to a check and so on (691).

On a similar vein, any proposition that describes a human action employing predicates of the special sciences fails to correspond to an equivalent description in the vocabulary of physics if that action is multiply realizable by distinct physical events. Take, for example, the proposition 'Jones will vote for the Democrats in the upcoming elections'. The event that falls under the special predicate 'vote' could be realized in many physical ways, such as writing something down in a piece of paper, raising one's hand, submitting an electronic form, making an utterance, and so on. That being the case, neither the special predicate 'vote' nor the whole proposition could be identified with a single physical description in virtue of which they would, individually or jointly, fall under some law of nature. As a consequence, reducing the special predicates included in some future descriptions of the world to the natural kind predicates figuring in the laws of nature does not seem to be a promising route for formulating a deductively valid *deterministic* argument.

*(ii) Freedom of will is at stake, not freedom of action*

Someone might object that what is at stake in the debate between compatibilists and incompatibilists is whether *freedom of will*, instead of *freedom of action*, is compatible with determinism. It may be true that the event described by the proposition 'Jones will vote for the Democrats in the upcoming elections' could be realized in many distinct physical events but, nevertheless, the event represented by the proposition 'Jones will *decide/intend* to vote for the Democrats in the upcoming elections' is uniquely realized in a single neurological event occurring in the human brain. Given what Van Inwagen says about the laws of chemistry -"Thus, a law about chemical valences is a law of physics in my sense, even if chemistry is not ultimately 'reducible' to physics" (187) - we can defensibly conclude that he would be willing to include the laws of neuroscience in the conjunction of all the natural laws, and further argue that we could reconstruct a deductively valid *deterministic* argument by reducing the terms describing future manifestations of the human will to the natural kind predicates of neuroscience.

Three responses to the above objection: first, a successful formulation of the

10

*deterministic* argument into a deductively valid argument requires more than the mere replacement of the proposition 'Jones will *decide/intend* to vote for the Democrats in the upcoming elections' with a physical description in the vocabulary of neuroscience. What is needed is a strict correspondence between the proposition and the natural kind predicates (=bound variables in scientific laws) appearing in the relevant law of neuroscience. Even if the event is uniquely realizable by a single neurological event in the brain, the description of the neurological event is not necessarily equivalent to a natural kind predicate of neuroscience, as in the example of 'monetary exchange' where its physical description ('signing one's name on a paycheck', say) does not correspond to any natural kind predicate of the laws of nature.

Second, Fodor argues that it's nomologically possible that creatures other than human beings (such as automata) could satisfy natural kind predicates in psychology without satisfying natural kind predicates in neuroscience. Given that every identity should be nomologically and counterfactually necessary, the nomologically possible realizability of psychological events by non-neurological events renders the identification of psychological natural kind predicates with neurological natural kind predicates impossible (692-3). Thus, the counterfactual possibility of the event 'Jones will *decide/intend* to vote for the Democrats in the upcoming elections' being multiply realizable by distinct physical events hinders the reduction of the proposition to the vocabulary of neuroscience and, as a result, subverts the reconstruction of the *deterministic* argument in a deductively valid way.

Third, the *deterministic* argument would be deductively valid if and only if every possible term appearing in the future description of the world was logically equivalent to some term appearing in the premises. As long as there are terms representing human actions ('vote', 'elections', etc.) that cannot be semantically reduced to the natural kind predicates of natural laws or the terms figuring in the physical description of the world in the remote past, the past and the conjunction of the laws of nature do not *logically* entail all future propositions. If determinism is necessarily true in the way suggested by the present interpretation, we should be able to formulate a deductively valid argument that *logically* entails in its conclusion all possible descriptions of the future state of the world and not just those that are relevant to the debate of *free will*.

11

*(iii) Why not include the laws of special sciences in the first premise?*

Someone might suggest including the laws of special sciences in the conjunction of all the laws of nature. By doing so, we could reduce some of the terms appearing in the conclusion of the *deterministic* argument to some of the terms appearing in its first premise. However, as previously said, Van Inwagen would not include the laws of special sciences - or, in his own terminology, the voluntaristic laws - in the conjunction of the laws of nature, for he thinks that determinism based on voluntaristic laws may be compatible with freedom of will. A *deterministic* argument including natural kind predicates of the special sciences in its first premise could be reconstructed in a deductively valid way, but it would not be compatible with the version of determinism that Van Inwagen considers to be crucial to the free will debate.

CONCLUSION

In this paper I have tried to provide two different interpretations of the first premise of Van Inwagen's modal argument for incompatibilism and show why the premise is false under each one of them. According to the first interpretation, the premise '□ ((P & L) → P)' is equivalent to the proposition '*Necessarily*, the past and the conjunction of the laws of nature entail the future', which could be refuted by showing that determinism is an empirically contingent view - tightly correlated with the success of scientific predictability - and so must be its logical consequences. According to the second interpretation, the premise '□ ((P & L) → P)' is equivalent to the proposition 'the past and the conjunction of the laws of nature *logically* entail the future', which would be true only if every proposition describing the world at a particular instant in the future was logically equivalent to some natural kind predicate of a natural science. The propositions describing future psychological events are irreducible to propositions representing natural kind predicates of neuroscience, and thus, the first premise is false. Assuming that I have provided an exhaustive list of interpretations, we can safely conclude that Van Inwagen's modal argument for incompatibilism fails to establish its conclusion because it presupposes a definition of determinism that is either incongruent with the prevailing current views or logically implausible.

# REFERENCES

1.Fodor, Jerry. "Special Sciences (or: The Disunity of Science as a Working Hypothesis)". In Michael Martin, and Lee C. McIntyre (eds*.), Readings in the Philosophy of Social Science*, pp. 687-699. Massachusetts: MIT Press, 1994.

2.Hoefer, Carl. "Causal Determinism". In *Stanford Encyclopedia of Philosophy*. 2010.

3.Van Inwagen, Peter. "The Incompatibility of Free Will and Determinism". In *Philosophical Studies 27*, pp. 185-199. 1975.

4.Van Inwagen, Peter. "A Modal Argument for Incompatibilism". In Paul Russell and Oisin Deery (eds.), *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*, pp. 117-125. New York: Oxford University Press, 2013.

5.Vihvelin, Kadri. "The problem distinguished: Is it Possible for Us to Have Free Will?". In *Causes, Laws and Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press, 2013.