Final Paper
PHILOS 136
Weiqi Yu

Fixing van Inwagen's Modal Argument for Incompatibility

1 Introduction

In *An Essay on Free Will*, van Inwagen introduced his famous modal argument for the incompatibility between free will and determinism, which involves a flavor of modal propositional logic with two new inference rules and one more modal operator he proposed, namely rule-α, rule-β, and the operator N. This argument has enabled us to think about these issues in a whole new way with the tools of modal logic. However, van Inwagen's modal argument, though inspiring, has been shown to have several issues. First, as he noticed himself, his argument involves a system of modal logic that is not fully constructed. It does not have formal semantics, and the soundness of the rules he proposed is not formally justified. What's worse, as pointed out by Warfield (2000), the conclusion yielded by van Inwagen's modal argument is "strictly weaker than the proper incompatibilist position (INC)."

In this paper, we will begin by briefly considering van Inwagen's original modal argument, which shall show us why its conclusion is strictly weaker than the proper one. After that, we will try to fully construct a system of modal propositional logic with N with well-defined syntax, semantics, and a deduction system slightly different from van Inwagen's. To have such a deduction system, we begin by considering a frame condition and see that this is a reasonable assumption to make in our reasoning about choice. Then we will present the rules of the deduction system, such that for frames satisfying this frame condition we will see how our deduction system proves to be sound. And within that system, we will formulate a proof of incompatibility in the stronger sense, but still in van Inwagen's style.

Thus, we get to show the proper incompatibilist position with proof in a sound deductive system (for frames of a specific but, at least from my expectation, very acceptable frame condition). Finally, we will consider possible objections to this argument.

2 What's Wrong with van Inwagen's Argument?

In the deduction system proposed by van Inwagen with rule-α ($\Box\varphi \vdash N\varphi$) and rule-$\beta$, the following can be proven:

$$\Box((P_0 \wedge L) \to P), NP_0, NL \vdash NP$$

However, as Warfield pointed out, this is weak: For us to get NP, it is required for all worlds in the frame to have the same law of nature L and some shared past state P_0. For example, consider the following model $\mathcal{M}_1$, in which solid lines stand for the accessibility relation of $\Box$, and dashed lines for that of N. L_0 and L_1 represent different conjunctions of the law of nature. $\Box((P_0 \wedge L_0) \to P)$ is true for all worlds in this model and is omitted for brevity.

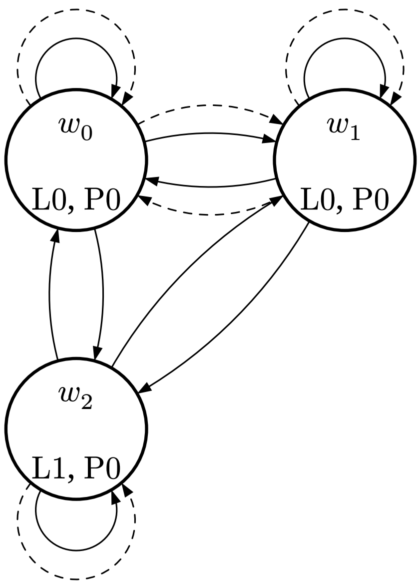

Figure 1: $\mathcal{M}_1 = (W, R_\Box, R_N, V)$

We can see that though $\mathcal{M}_1, w_0 \Vdash NL_0 \wedge NP_0 \wedge \Box((P_0 \wedge L_0) \to P)$, van Inwagen's conclusion does not assure that $\mathcal{M}_1, w_0 \Vdash NP$ even if his deduction system is sound, since it is not the case that $(W, R_\Box, R_N) \Vdash NL_0$.

In the rest of this paper, we will attempt to formulate a formal and stronger version of van Inwagen's argument, with the proper incompatibilist position (INC) proven.

3 The Modal Operator N, Formally Defined

3.1 Syntax

Definition 3.1.1:

The syntax for the formula in this logic can be defined in BNF form as follows:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \varphi \leftrightarrow \varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \Box\varphi \mid \Diamond\varphi \mid N\varphi$$

3.2 Semantics

Van Inwagen frames the meaning of his modal operator N as follows: for any proposition p, Np means "p and no one has, or ever had, about whether p".[1]

To capture this, we first define the accessibility relation for N as follows:

Definition 3.2.1:

$R_N \subset W \times W$ is a binary relation such that wR_Nw' iff w' is can be reached from w with and/or without the choice of someone in w.

With R_N, the semantics of N can be defined formally as follows:

[1] van Inwagen, An Essay on Free Will, p. 93

Definition 3.2.2:

For a Kripke model $\mathcal{M} = (W, R_\Box, R_N, V)$ where W is the set of worlds, $R_\Box$ is the accessibility relation for the broad logical necessity modal operator $\Box$, R_N is the accessibility relation for N, and V is a valuation:

$$\mathcal{M}, w \Vdash N\varphi \text{ iff } \forall w' \in R_N[w].\mathcal{M}, w' \Vdash \varphi$$

I believe this semantics captures van Inwagen's conception of N, because:

1. By the definition of R_N, we have $\forall w \in W.w \in R_N[w]$. Thus this formal semantics of N defined here captures the first part of van Inwagen's formulation (before the word "and"), i.e., $w \Vdash N\varphi \text{ implies } w \Vdash \varphi$.
2. For the second part of his formulation (after the word "and"), observe that when all worlds than can be reached with and/or without the choice of someone in w force φ, nobody has or ever had any choice about whether φ.

4 Frame Condition N

Definition 4.1:

For a Kripke frame $(W, R_\Box, R_N)$ where W is the set of worlds, $R_\Box$ is the accessibility relation for $\Box$, and R_N is the accessibility relation for N, the Frame Condition N is

$$\forall w \in W.R_N[w] \subset R_\Box[w]$$

A Kripke frame satisfying Frame Condition N is called an N-Frame in this paper.

4.1 Can we accept this frame condition?

Since $\Box$ represents broad logical necessity, for any world w it should be understood that $R_\Box[w]$ is the set of worlds possible in the broad logical sense. If that is the case, to assert Frame Condition N would be no more than to assert that for any world

w' that can be reached from w with or without the choice of someone, w' must be possible in broad logical sense, which is an acceptable assumption to make.

5 From Axiom-N to System-N

5.1 Soundness of Axiom-N

Theorem 5.1.1 (Axiom-N):

$$\Vdash_{\boldsymbol{N}} \Box\varphi \to N\varphi$$

Proof: Let $(W, R_\Box, R_N)$ be an N-Frame, where W is the set of worlds, $R_\Box$ is the accessibility relation for $\Box$, and R_N is the accessibility relation for N. Let φ be any formula of modal propositional logic. We will show that $\forall w \in W.w \Vdash \Box\varphi \to N\varphi$. To do so, let w be a world in W and assume that $w \Vdash \Box\varphi$. We will show that $w \Vdash N\varphi$. Note that by the semantics of the $\Box$ operator, $\forall w' \in R_\Box[w].w' \Vdash \varphi$. By Definition 4.1, $R_N[w] \subset R_\Box[w]$. Thus $\forall w' \in R_N[w].w' \Vdash \varphi$. By the semantics of the N operator, $w \Vdash N\varphi$. ■

5.2 Correspondence

Theorem 5.2.1:

For any Kripke frame $(W, R_\Box, R_N)$ where W is the set of worlds, $R_\Box$ is the accessibility relation for $\Box$, and R_N is the accessibility relation for N, if $(W, R_\Box, R_N) \Vdash \Box\varphi \to N\varphi$, then $\forall w \in W.R_N[w] \subset R_\Box[w]$.

Proof: Let $(W, R_\Box, R_N)$ be a Kripke frame where W is the set of worlds, $R_\Box$ is the accessibility relation for $\Box$, and R_N is the accessibility relation for N. Assume $(W, R_\Box, R_N) \Vdash \Box\varphi \to N\varphi$. Let w be any world in W, we will show that $R_N[w] \subset R_\Box[w]$. Let $V(p) = R_\Box[w]$ be a valuation and $\mathcal{M} = (W, R_\Box, R_N, V)$ be a Kripke model. By the assumption, $\mathcal{M}, w \Vdash \Box p \to Np$. Since $V(p) = R_\Box[w]$, $\forall w' \in R_\Box[w].\mathcal{M}, w' \Vdash p$. So, by the semantics of $\Box$, we have $\mathcal{M}, w \Vdash \Box p$. Then, by the semantics of the conditional, we have $\mathcal{M}, w \Vdash Np$. Thus by the semantics of N, we have $\forall w' \in R_N[w].\mathcal{M}, w' \Vdash p$. That is to say, $R_N[w] \subset V(p)$. Therefore by the definition of $V(p)$, $R_N[w] \subset R_\Box[w]$. ■

Theorem 5.2.2 (Correspondence):

$$(W, R_\Box, R_N) \Vdash \Box\varphi \to N\varphi \text{ iff } \forall w \in W.R_N[w] \subset R_\Box[w]$$

Proof: Two directions of this biconditional are shown by Theorem 5.1.1 and Theorem 5.2.1. ■

With the correspondence theorem, we can see how Frame Condition N is not only a sufficient condition for Axiom-N, but also a necessary one.

5.3 System-N

5.3.1 Definition

Definition 5.3.1.1:

System-N is the deduction system formed by adding Axiom-N to System-K:

$\varphi \vdash \Box\varphi$	Nec
$\varphi, \varphi \to \psi \vdash \psi$	Modus Ponens
$\Box\varphi \to N\varphi$	Axiom-N Schema
$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	Axiom-K Schema for $\Box$
$N(\varphi \to \psi) \to (N\varphi \to N\psi)$	Axiom-K Schema for N
Any tautology of classical propositional logic is an axiom	PL

5.3.2 Soundness

Theorem 5.3.2.1:

For any formula of modal logic φ,

$$\text{If} \vdash_N \varphi \text{ , then} \Vdash_N \varphi.$$

Proof: With Theorem 5.1.1, the soundness of System-N in N-Frames can be shown similarly to the proof of that of System-K. Consider the proof for the Soundness Theorem of System-K (as in Lecture Note 9.3). By adding Axiom-N to the base case and the first subcase of the induction step, and repeating subcase 3 but for the modal operator N instead, we get proof for the soundness of System-N. ■

6 A Stronger Argument, in System N

Theorem 6.1:

$$\vdash_{\boldsymbol{N}} \Box((\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to NP)$$

Proof:

1.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to \Box((P_0 \wedge L) \to P)$	PL
2.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to NP_0$	PL
3.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to NL$	PL
4.$((P_0 \wedge L) \to P) \to (P_0 \to (L \to P))$	PL
5.$\Box(((P_0 \wedge L) \to P) \to (P_0 \to (L \to P)))$	Nec 4
6.$\Box(((P_0 \wedge L) \to P) \to (P_0 \to (L \to P))) \to (\Box((P_0 \wedge L) \to P) \to \Box(P_0 \to (L \to P)))$	$K_\Box$
7.$\Box((P_0 \wedge L) \to P) \to \Box(P_0 \to (L \to P))$	MP 5, 6
8.$\Box(P_0 \to (L \to P)) \to N(P_0 \to (L \to P))$	N
9.$N(P_0 \to (L \to P)) \to (NP_0 \to N(L \to P))$	K_N
10.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to (NP_0 \to N(L \to P))$	PL 1, 7, 8, 9
11.$(\varphi \to \psi_1) \to ((\varphi \to (\psi_1 \to \psi_2)) \to (\varphi \to \psi_2))$	PL
12.$((\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to (NP_0 \to N(L \to P)))$	
$\to ((\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to N(L \to P))$	MP 2, 11
13.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to N(L \to P)$	PL 10, 12
14.$N(L \to P) \to (NL \to NP)$	K_N
15.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to (NL \to NP)$	PL 13, 14
16.$((\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to (NL \to NP))$	
$\to ((\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to NP)$	MP 3, 11
17.$(\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to NP$	MP 15, 16
18.$\Box((\Box((P_0 \wedge L) \to P) \wedge NP_0 \wedge NL) \to NP)$	Nec 17

Note that the φ and ψ used in this proof are merely for space saving. It shall be trivial to see how they can be substituted by concrete instances of formulae of modal logic for the tautology to be used in different contexts of the proof.

■

Corollary 6.1.1:

By Theorem 5.3.2.1 and Theorem 6.1,

$$\Vdash_{\boldsymbol{N}} \Box((\Box((P_0 \wedge L) \rightarrow P) \wedge NP_0 \wedge NL) \rightarrow NP)$$

i.e., in N-Frames, necessarily, if determinism is true, and no one has or ever had any choice about the truth of P_0 and L, then no one has or ever had any choice about P; i.e., in N-Frames, necessarily, if determinism is true then there is no freedom. Thus, Warfield's INC is proven in N-Frames: As long as one asserts that nobody's action can or ever could render their world impossible in the broad logical sense, one would have to agree with the truth of INC.

7 Possible Objections

There are two objections that I can think of:

1. This proof still forces a specific set of natural laws L and a shared past state P_0.
2. Determinism doesn't require the strict conditional $\Box((P_0 \wedge L) \rightarrow P)$; instead, something like a Lewisian counterfactual $(P_0 \wedge L) \Box\!\!\rightarrow P$ would be enough. Thus, even though the proof is valid, it does not really show incompatibility.

For (2), I do not have a good response in mind for now. If you believe that determinism doesn't require the strict conditional, you can consider our argument to be one with a weaker conclusion for the incompatibility between free will and a specific kind of determinism. But for (1), notice that for each world w with a set of law of nature L_w and distant past P_{0w}, we can construct the same proof with L

replaced by L_w and P_0 replaced by P_{0w}. The conjunction of the conclusion of all these proofs should fully address the concern of (1).