

Impossible Worlds and Logical Omniscience: An Impossibility Result*

Jens Christian Bjerring

Forthcoming in *Synthese*

Abstract

In this paper, I investigate whether we can use a world-involving framework to model the epistemic states of non-ideal agents. The standard possible-world framework falters in this respect because of a commitment to logical omniscience. A familiar attempt to overcome this problem centers around the use of impossible worlds where the truths of logic can be false. As we shall see, if we admit impossible worlds where “anything goes” in modal space, it is easy to model extremely non-ideal agents that are incapable of performing even the most elementary logical deductions. A much harder, and considerably less investigated challenge is to ensure that the resulting modal space can also be used to model moderately ideal agents that are not logically omniscient but nevertheless logically competent. Intuitively, while such agents may fail to rule out subtly impossible worlds that verify complex logical falsehoods, they are nevertheless able to rule out blatantly impossible worlds that verify obvious logical falsehoods. To model moderately ideal agents, I argue, the job is to construct a modal space that contains only possible and non-trivially impossible worlds where it is not the case that “anything goes”. But I prove that it is impossible to develop an impossible-world framework that can do this job and that satisfies certain standard conditions. Effectively, I show that attempts to model moderately ideal agents in a world-involving framework collapse to modeling either logical omniscient agents, or extremely non-ideal agents.

*I would like to thank especially David Chalmers, Alan Hájek, Mark Jago, Wolfgang Schwarz, and Weng Hong Tang for invaluable discussion of my ANU dissertation on which this paper is based. And I would like to thank David Chalmers, Julien Dutant, Daniel Nolan, Nikolaj J. L. L. Pedersen, Wolfgang Schwarz, Weng Hong Tang, two anonymous referees, audience at the 2010 Copenhagen-NIP Formal Epistemology Workshop, audience in the hyperintensionality symposium at the 14th Congress of Logic, Methodology and Philosophy of Science, and the Research Group for Analytic Philosophy at Aarhus University for further helpful discussion of this paper.

1 Introduction

In a letter to Graham Priest, David Lewis writes:

I'm increasingly convinced that I can and do reason about impossible situations. [...] But I don't really understand how that works. Paraconsistent logic as developed by you [Priest] and your allies is clear enough, but I find it a bit off the topic. For it allows (a limited amount of) reasoning about *blatantly* impossible situations. Whereas what I find myself doing is reasoning about *subtly* impossible situations, and rejecting suppositions that lead fairly to blatant impossibilities.

[...] I agree with you [Priest] about the many uses to which we could put make-believedly possible impossibilities, if we are willing to use them. The trouble is that all these uses seem to require a distinction between the subtle ones and the blatant ones [...] and that's just what I don't understand. (Lewis 2004, pp. 176-177).

Intuitively, mathematical falsehoods such as '0=1', conceptual or analytic falsehoods such as 'The bike is red and blue all over', and logical falsehoods such as 'It rains and it does not rain' are blatantly impossible or inconsistent. In contrast, mathematical falsehoods such as 'There are integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$ ', and logical falsehoods such as 'One day there will be recursive computers that can prove any mathematical statement that is true' are subtly impossible or inconsistent.

Blatant inconsistencies are the kinds of inconsistencies that ordinary, reflective reasoners like Lewis can easily rule out independently of what empirical information they might happen to have. For instance, if I were to reflect on an inconsistency such as 'It rains and it does not rain', I do not have to consult the world to make sure that it is always false. Rather, I would immediately deem such blatant inconsistencies false by a cognitively trivial chain of a priori reasoning.

In contrast, subtle inconsistencies are the kinds of inconsistencies that ordinary, reflective reasoners like Lewis cannot easily rule out unless they possess the relevant empirical information. Although I have excellent empirical reasons—acquired through testimony say—to hold that Fermat's Last Theorem is true, I still cannot justify the theorem using solely a priori reasoning. Intuitively, if I were to suspend all my empirical beliefs and consider whether Fermat's Last Theorem is true, I would not reach a verdict. For I would not be able to emulate the highly non-trivial chain of a priori reasoning that demonstrates that an inconsistency such as 'There are integers $a, b, c, n > 2$ such that $a^n + b^n = c^n$ ' is false.

In this paper, I will investigate whether we can use a world-involving framework to model the epistemic states of non-ideal agents like Lewis. Such agents, I will assume, have bounded but non-trivial cognitive and computational resources available for rational a priori reasoning. And this, I will assume, explains why such agents are always in a position to easily rule out blatantly impossible situations, but often not in a position to rule out subtly impossible ones. More specifically, I will investigate whether we can use an impossible-world framework to model the epistemic states of *moderately ideal agents* that always engage in the kind of logical reasoning that ordinary, reflective reasoners like you and I can easily, and often do engage in. The following impossibility theorem will show the difficulties involved in developing an impossible-world framework that can do this job:

(Result) There is no modal space such that:

- (R₁) there are impossible worlds;
- (R₂) there are no partial worlds;
- (R₃) there are no blatantly impossible worlds.

To establish (Result), I proceed as follows.

In section 2 I provide background information. In section 3 I make the notion of blatant inconsistencies precise. In sections 4 and 5 I investigate two constructions of modal space that will help explain and motivate (R₁) through (R₃) in (Result). In section 4 I construct *Logically Ideal Space*. This space bears resemblance to a standard possible-world framework for reasoning about knowledge. The space is *logically ideal* because it applies to *ideal agents* that are logically omniscient. The construction is based on a standard maximal-consistency conception of worlds according to which, roughly, a world corresponds to a maximally specific, logically consistent description of a way the world might be. In addition to introducing the basic technical machinery that I will use to prove (Result), section 4 helps explain (R₁) and (R₂). When we know which worlds model logically omniscient agents, we will also know which worlds model logically non-omniscient agents: namely logically inconsistent or impossible worlds. And when we know what maximal worlds are, we will also know what partial worlds are: namely worlds that are not maximal. In section 5 I construct *Logically Extreme Space*. This space bears resemblance to a standard impossible-world framework for reasoning about knowledge. The space is *logically extreme* because it applies to *extremely non-ideal agents* that are incapable of even elementary logical reasoning.

Logically Extreme Space will contain an abundance of blatantly impossible worlds—worlds, that is, that verify blatant logical inconsistencies. But to model the class of *moderately ideal agents* that are logically competent but not logically omniscient, I argue that we need a construction of a *logically non-trivial space* that contains no blatantly impossible worlds. Section 5 thus helps motivate condition (R₃) and (R₂)—although I will not spend much too time on (R₂) in this paper. In section 6 I prove (Result). This will show that we need blatantly impossible worlds in modal space if we want to avoid logical omniscience. In turn, this will show that attempts to construct a logically non-trivial space collapse to modeling either extremely non-ideal agents—in case modal space contains blatantly impossible worlds—or logically omniscient agents—in case modal space contains no blatantly impossible worlds. In section 7 I conclude.

2 Background

A standard possible-world framework for reasoning about knowledge can be said to have two core features. First, knowledge is modeled in terms of quantification over worlds. When an agent knows that Canberra is the capital of Australia, then Canberra is the capital of Australia in all worlds that are epistemically possible for the agent. Intuitively, a world corresponds to a maximally specific hypothesis about a way things might be. And roughly, we can say that a world is epistemically possible for an agent when the world is not ruled out by anything the agent knows. Second, knowledge aggregation is modeled in terms of elimination of worlds. When an agent comes to know that Canberra is the capital of Australia, all worlds where Canberra is not the capital of Australia are ruled out as epistemically possible for the agent.

We can state this world-involving representation of intentional notions such as belief and knowledge as follows:¹

(**Know**) An agent a knows a sentence A iff A is true at all worlds that are epistemically possible for a .

To investigate whether we can use a world-involving framework to model the epistemic states of moderately ideal agents, we can investigate whether (Know) can be made plausible for such agents.

¹Throughout the paper I will talk about knowledge rather than belief, but the main insights apply mutatis mutandis to beliefs.

The world-involving framework presupposes a space of all worlds that I will call *modal space*. To make (Know) plausible for moderately ideal agents, I will take the basic challenge to involve constructing a modal space in which every world *may* remain epistemically possible for such agents. Moderately ideal agents have non-trivial but limited cognitive capacities available for a priori logical reasoning. Because of this, they may fail to know that subtle logical inconsistencies are false, but never that blatant logical inconsistencies are false. As such, subtly impossible worlds—worlds that verify subtle logical inconsistencies—may remain epistemically possible for these agents, whereas blatantly impossible worlds may not. The task is then to construct a modal space that contains subtly impossible worlds, but no blatantly impossible worlds.

As I have set up the basic challenge, blatantly impossible worlds should be *excluded* from a logically non-trivial space. On a slightly different way of setting up the challenge, however, we allow that modal space contains blatantly impossible worlds, but aim to ensure that no such worlds remain epistemically possible for moderately ideal agents. Understood in this way, there are worlds in modal space that could never be epistemically possible for moderately ideal agents. Although nothing of substance will change if we adopt this alternative formulation of the basic challenge—in particular, (Result) will hold in either case—I prefer the more direct route through a construction of modal space itself. The main reason is this. Regardless of what circumstantial empirical information moderately ideal agents happen to have, blatantly impossible worlds are *always* epistemically impossible for them. So to model these kinds of agents, blatantly impossible worlds will not play a role, and consequently they need not be included in a logically non-trivial space.² But I should stress that the main conclusions remain unaffected by this choice of strategy.

I represent the objects of knowledge—whatever they are—by *sentences*. Although this is a slightly non-standard representation, it is natural to invoke sentences for much of the constructive work that I want to do in this paper. For simplicity, I take these sentences to be declarative sentence *types* in a language \mathcal{L} with no context-dependent expressions. This will allow us to represent the epistemic states of agents as having definite contents. As a rough guide, we can think of \mathcal{L} as a version of English—

²Of course, blatantly impossible worlds may play a role in modeling different kinds of agents. But, as will become clear in the discussion in section 5, it is easy to think of a logically non-trivial space for moderately ideal agents in terms of a subspace of an all-inclusive space containing arbitrary blatantly impossible worlds.

sufficiently enriched with mathematical and logical expressions—in which all context-dependent expressions have been replaced by corresponding context-independent expressions. Further, I assume that \mathcal{L} has two symbols \neg and \rightarrow that play the same inferential roles as classical negation and material implication. We can then treat the other standard connectives as shorthand for the familiar definitions in terms of \neg and \rightarrow . Henceforth, we should understand all unqualified talk of sentences as involving talk about sentence types in \mathcal{L} . For obvious reasons, however, I will continue to phrase examples in plain English.³

In its abstract form, the world-involving framework is compatible with different ways of conceiving of worlds in modal space. Standardly, worlds are identified with *possible worlds*, where possible worlds are standardly identified with maximal, consistent sets of propositions or interpreted sentences in some world-making language. In this linguistic ersatz tradition, we find Robert Adams’ complete consistent sets of propositions (Adams 1974), Rudolf Carnap’s state descriptions (Carnap 1947), David Chalmers’ equivalence classes of epistemically complete sentences (Chalmers 2011), and Jaakko Hintikka’s and Richard Jeffrey’s complete consistent novels (Hintikka 1962, Hintikka 1969, Jeffrey 1983). I will follow this linguistic ersatz tradition and identify worlds with sets of sentences in a world-making language. In particular, I will identify the world-making language with the language \mathcal{L} . Since I assume that \mathcal{L} has no context-dependent expressions, \mathcal{L} constitutes a suitable “disambiguated and precise” language for constructing a space of worlds that we can use to model agents whose epistemic states are already represented by sentences in \mathcal{L} .⁴

3 Blatant and Subtle Inconsistencies

We seem to have a good pre-theoretical grasp of which inconsistencies count as blatantly inconsistent: namely, the ones that we can rule out easily, effortlessly, or

³It is worth noticing that one might identify the objects of knowledge with *propositions*, where these are understood as primitive abstract entities that can capture arbitrarily fine-grained cognitive and epistemological differences in non-ideal reasoning. If we understand propositions in this way, it is open for the reader to translate everything I say about sentences in \mathcal{L} to propositions of this fine-grained sort.

⁴Cf. Lewis 1986, p. 142. The current approach resembles the approach taken in Jeffrey 1983. Much like I do, Jeffrey conceives of the world-making language—in which he also represents the beliefs and utilities of agents—as “idealized in the sense that the (declarative) sentences of that language have fixed [...] truth-values, independent of the contexts of their utterance.” (Jeffrey 1983, p. 208).

immediately. But it is useful to work with a more precise notion in what follows.⁵ Generally, I will restrict my attention to *propositional logical* blatant inconsistencies. So even if a modal space contains worlds that verify blatant inconsistencies such as ‘My bike is red and blue all over’, ‘ $2 + 2 = 5$ ’, and ‘ $\exists x(Fx \wedge \neg Fx)$ ’, it may still be successful as a space for moderately ideal agents that are characterized solely in terms of their capacities for limited propositional logical reasoning. Whenever I talk about (logical) blatant and subtle inconsistencies—and (logically) possible and impossible worlds—I will henceforth omit the reference to *logical* blatant and subtle inconsistencies. Likewise, whenever I talk about (propositional) logical truths, (propositional) logical entailments, and the like, I will henceforth omit the reference to *propositional* logical truths and entailments.

The distinction between blatantly and subtly impossible worlds that we are looking for is thus a distinction between logically inconsistent worlds that can be easily, effortlessly, or immediately ruled out using logic, and ones that cannot. On the face of it, we should expect that we can draw such a distinction—although perhaps not a sharp one. Consider a parallel with the relative difficulty of proofs. Whereas some formulas in propositional logic can be disproved using just a few applications of the standard inference rules, other formulas cannot. It seems that something similar should apply in the case of inconsistent worlds. Whereas some inconsistent worlds—the blatantly inconsistent ones—verify sentences that can be disproved using just a few applications of the inference rules, other inconsistent worlds—the subtly inconsistent ones—do not.

What counts as a blatant inconsistency is dependent on the kinds of agents that we are interested in—just as what counts as an easy proof is agent-dependent. Whereas certain complex logical falsehoods may count as blatantly inconsistent for highly well-trained logicians, they need not do so for first-year logic students. Since my main focus here is on moderately ideal agents, I will use a precise characterization of these agents to get clear on the notion of blatant inconsistencies.

Moderately ideal agents, on my view, have limited but non-trivial cognitive and computational capacities for easy (propositional) logical reasoning. I will take this to mean that these agents know and can apply the inference rules of propositional logic to extend their knowledge. Yet, since they have only limited cognitive resources, I will say that they can only apply these inference rules a limited number of times. So,

⁵Thanks to an anonymous referee for stressing this point.

in line with Jago, I have a resource-bounded conception of rationality in mind:

A coherent, deductively closed set of beliefs is an ideal of rational enquiry, for example, yet an agent can be deemed rational if it has the ability to reason in accordance with certain logical rules and it deploys those abilities as well as the cognitive resources to hand allow. Failures of closure within an agent's belief set may be due to a failure of rationality but they may also be due to a lack of cognitive resources. (Jago 2009, p. 333.)

We can take the relevant inference rules—with three minor caveats below—to be the inference rules of a standard natural deduction system similar to the ones found in Bostock 1997 or Copi 1979.

More precisely, I will say that a moderately ideal agent is an agent that never fails to know any sentence that she can infer from what she already knows by applying a given inference rule *once*—by performing 1 step of logical reasoning. For instance, suppose that a moderately ideal agent knows $(A \wedge B)$ and knows how to apply conjunction elimination. On the current view, this agent would then never fail to know, say, A . For she can infer A from her knowledge by 1 step of logical reasoning involving conjunction elimination. In contrast, suppose a moderately ideal agent knows $((A \wedge B) \wedge C)$. Although this agent will never fail to know, say, C , she may well fail to know A or B . For she can neither infer A nor B from her knowledge by 1 step of logical reasoning involving conjunction elimination—it takes two applications of conjunction elimination to do so.

The idea that moderately ideal agents always extend their knowledge by 1 step of logical reasoning encodes the intuitive idea of *easy*, *effortless*, or *immediate* logical reasoning. The framework that I will be developing describes the epistemic states of agents at times. These states, as I will understand them, include not only the sentences that agents know at t , but also the sentences that they can infer from this stock of knowledge by applying an inference rule once. Reasoning with rules for the connectives takes time. So when I characterize the epistemic state of an agent at a time, this time unit is intuitively stretched out to include the brief time that it takes to apply an inference rule. In a sense, this makes the framework dynamic. But the dynamics should be understood at a very coarse level of grain that abstracts away from whatever happens in the interval where an inference rule is applied.

This 1-step characterization of moderately ideal agents presupposes only a weak requirement of rationality. One might worry that this requirement is too weak. For

example, instances of $(A \wedge \neg A)$ seem to count as blatantly inconsistent. Since we want to ensure that moderately ideal agents *always* rule out blatant inconsistencies, the 1-step characterization then seems insufficient. For supposing that “ruling out”, in the current context, means “inferring an explicit contradiction A and $\neg A$ ”, it takes normally two applications of conjunction elimination to infer A and $\neg A$ from $(A \wedge \neg A)$. If so, the 1-step characterization will not do the required work.

If inconsistencies of the form $(A \wedge \neg A)$ are indeed blatant, there are two natural ways to ensure that moderately ideal agents always rule them out. Either we introduce a slightly non-standard rule of conjunction elimination that allows an agent to infer both A and B from a conjunction $(A \wedge B)$ in 1 step. Or we say that moderately ideal agents may apply the inference rules at most n times, for some sufficiently small $n > 1$. For my purposes, it would suffice to let $n = 2$, but other values for n are of course possible.

Either fix will work for me. To make things run as smoothly as possible, however, I will work with the 1-step characterization of moderately ideal agents and make three specific stipulations about the inference rules—these regard the three caveats mentioned above. First, agents can reason using a rule for conjunction elimination that permits inferring in 1 step both A and B from a conjunction $(A \wedge B)$. This stipulation ensures that moderately ideal agents always rule out blatant inconsistencies of the form $(A \wedge \neg A)$. Second, agents can reason using a rule (R-NC) for negated conditionals that permits inferring in 1 step A and $\neg B$ from $\neg(A \rightarrow B)$. This stipulation will allow me to keep the proof of (Result) clean and simple. As I think of it, (R-NC) is a natural deduction version of the conjunction elimination rule from a standard tableaux system, the latter of which allows us to infer A and $\neg B$ on a branch, if $\neg(A \rightarrow B)$ already occurs on the branch. (R-NC) is not an unreasonable rule for agents that are minimally logically competent, and it is justified by appeal to (implicit) knowledge of the truth-table for the conditional: a conditional is false only when A is true and B false. Third, agents can rule out sentence pairs of the form $\{A, \neg A\}$ non-inferentially. This stipulation ensures that we do not need to bother about the trivial inference rule that permits inferring in 1 step A from any set Γ such that $A \in \Gamma$.

Let me stress that nothing of importance hangs on these stipulations about the inference rules. By adopting the 1-step characterization of moderately ideal agents, it becomes easy to present and appreciate (Result). But, as we shall see, the core

idea behind (Result) is not lost if we do not make these stipulations. And, if need be, I am happy to adopt a characterization of moderately ideal agents—and, in turn, of blatant inconsistencies—that involves reference to n steps of logical reasoning, for some sufficiently small $n > 1$.

Given the 1-step characterization of moderately ideal agents, we can then give the following definition of what it means for something to be a blatant inconsistency:

(Blatant) A sentence A or a set Γ of sentences is blatantly inconsistent iff a contradiction $\{A, \neg A\}$ can be inferred from A or Γ by use of at most 1 application of the inference rules; otherwise, if inconsistent, A or Γ is subtly inconsistent.

Consider the inconsistent set $\Delta = \{A, (A \rightarrow B), \neg B\}$. By 1 application of modus ponens, B can be inferred from Δ . Since $\neg B$ is already in Δ , Δ can be ruled out in 1 step and is hence blatantly inconsistent—remember that agents can rule out non-inferentially sentence pairs of the form $\{A, \neg A\}$. In contrast, consider the inconsistent set $\Theta = \{A, B, (A \rightarrow (B \rightarrow C)), \neg C\}$. Since it takes two applications of modus ponens to infer a contradiction C and $\neg C$ from Θ , Θ is subtly inconsistent.

The 1-step characterizations of moderately ideal agents and blatant inconsistencies serve my theoretical purposes. For they will allow me to say that blatantly impossible worlds—worlds that verify a blatant inconsistency—are the kinds of worlds that moderately ideal agents always rule out as epistemically possible. That is, they will allow me to say that blatantly impossible worlds never remain epistemically possible for moderately ideal agents. In conjunction with (Know), this will ensure that moderately ideal agents always know that blatant inconsistencies are false.

But (Blatant) also captures core intuitions about blatant inconsistencies. Blatant inconsistencies are the kinds of inconsistencies that ordinary, reflective reasoners can easily, effortlessly, or immediately rule out by simple reasoning. Insofar as ordinary reasoners are capable of basic logical reasoning, (Blatant) captures this intuition. Blatant inconsistencies are the kinds of inconsistencies that ordinary, reflective reasoners can easily rule out independently of what empirical information they might happen to have. Insofar as logical reasoning is a priori in nature, (Blatant) captures this intuition as well.

On the current definition, an instance of a blatant inconsistency such as $(A \wedge \neg A)$ can be of arbitrary finite length. In some sense, this might make us doubt the intuitive plausibility of (Blatant). Yet, I do not think we should doubt (Blatant) for reasons to

w as follows:

(Truth) A sentence A is *true* at a world w iff $A \in w$.

(Falsity) A sentence A is *false* at a world w iff $A \notin w$.

If A is true at w , I will also say that w *verifies* A . If A is false at w , I will also say that w *falsifies* A . Third, I require that worlds are minimally closed in the following sense:

(Min-Clo) A world is *minimally closed* iff for all sentences A , $A \in w$ iff $\neg A \notin w$.

Finally, I require that worlds are consistent:

(Con) A set Γ of sentences is *consistent* iff Γ is satisfiable, where Γ is *satisfiable* iff there is a propositional evaluation that makes all sentences in Γ true.

The relevant evaluation should be understood as a *classical* propositional evaluation. If we let \mathcal{I} be an *interpretation* function that assigns either *true* or *false*, but not both, to each atomic sentence A in \mathcal{L} , then ν is the following classical evaluation function:

$(\nu\mathcal{I})$ If A is an atomic sentence in \mathcal{L} , then $\nu(A) = \mathcal{I}(A)$.

$(\nu\neg)$ $\nu(\neg A) = T$ iff $\nu(A) = F$.

$(\nu\rightarrow)$ $\nu(A \rightarrow B) = T$ iff $\nu(A) = F$ or $\nu(B) = T$.

The semantic clauses for \neg and \rightarrow reflect the idea that \mathcal{L} has symbols that play the same inferential roles as classical negation and material implication.

Given this, it is a familiar fact that (Sat)—which will play a central role in the proof of (Result) in section 6—can be proved:⁶

⁶The proof of (Sat) goes as follows:

Proof Let Γ be any set of sentences that satisfies (i) and (ii). We want to show that there is a propositional evaluation function ν that makes each $A \in \Gamma$ true. To this end, we stipulate an interpretation \mathcal{I} such that for all atomic A :

$$\mathcal{I}(A) = T \text{ iff } A \in \Gamma.$$

$$\mathcal{I}(A) = F \text{ iff } A \notin \Gamma.$$

This is a possible stipulation because \mathcal{I} cannot assign both T and F to any atomic A . We then need to show that every sentence in Γ is true under this interpretation. I do this by induction on the length of a sentence, where the length of a sentence is given by the number

(**Sat**) Any set of sentences Γ that satisfies the following two conditions is satisfiable:

- (i) $A \in \Gamma$ iff $\neg A \notin \Gamma$.
- (ii) $(A \rightarrow B) \in \Gamma$ iff $A \notin \Gamma$ or $B \in \Gamma$.

(Sat) allows us to say that every set of sentences—and hence every world—that satisfies (i) and (ii) is consistent, and that every inconsistent set of sentences either does not satisfy (i) or does not satisfy (ii).

From the principles (Min-Clo), (Truth), and (Falsity), we can also immediately establish that worlds satisfy:

(**Maximality**) For all sentences A and worlds w , either A is true at w or $\neg A$ is true at w .⁷

When a world w —or a set of sentences more generally—obeys (Maximality), I will also say that w is *maximal*. Let Logically Ideal Space be the class W_I of worlds that are maximal and consistent.

Intuitively, (Maximality) ensures that worlds can deliver an answer to each question that could in principle be asked about a particular way the world might be. Following Jeffrey, the question “Is A true in the maximally specific hypothesis that

of symbols it contains—for instance, $\neg A$ is longer than A , and $(A \rightarrow B)$ is longer than A and B :

Base case: Assume for atomic A that $A \in \Gamma$. We want to show that $\nu(A) = T$. We get the result immediately. By definition of \mathcal{I} , $A \in \Gamma$ iff $\mathcal{I}(A) = T$. By $(\nu\mathcal{I})$, $\mathcal{I}(A) = T$ iff $\nu(A) = T$. So $A \in \Gamma$ iff $\nu(A) = T$. So $\nu(A) = T$.

Inductive step: Assume for the induction hypothesis that every sentence in Γ that is shorter than $\neg A$ and $(A \rightarrow B)$ is true under the evaluation ν based on \mathcal{I} . We want to show that if $\neg A \in \Gamma$, then $\nu(\neg A) = T$, and if $(A \rightarrow B) \in \Gamma$, then $\nu(A \rightarrow B) = T$. There are two cases to consider:

Case 1: Assume $\neg A \in \Gamma$. By (i), $\neg A \in \Gamma$ iff $A \notin \Gamma$. By induction hypothesis, $A \notin \Gamma$ iff $\nu(A) = F$. By $(\nu\neg)$, $\nu(A) = F$ iff $\nu(\neg A) = T$. So $\nu(\neg A) = T$.

Case 2: Assume $(A \rightarrow B) \in \Gamma$. By (ii), $(A \rightarrow B) \in \Gamma$ iff $A \notin \Gamma$ or $B \in \Gamma$. By induction hypothesis, $A \notin \Gamma$ or $B \in \Gamma$ iff $\nu(A) = F$ or $\nu(B) = T$. By $(\nu\rightarrow)$, $\nu(A \rightarrow B) = T$ iff $\nu(A) = F$ or $\nu(B) = T$. In either case, $\nu(A \rightarrow B) = T$.

This completes the induction and thus the proof of (Sat). ■

⁷For suppose A is false at a world w . By (Falsity), $A \notin w$. Suppose, for reductio, that $\neg A$ is also false at w . By (Falsity), $\neg A \notin w$. By (Min-Clo), $A \in w$. Contradiction. So if A is false at w , then $\neg A$ is true at w . A similar line of reasoning applies in the case where $\neg A$ is false at w .

world w describes?” will be answered yes if $A \in w$ and no if $\neg A \in w$. (Jeffrey 1983, pp. 208-209). With a view to condition (R₂) in (Result), we can then say that a *partial* world is a world that fails to be maximal. That is, a partial world is a world at which neither A nor $\neg A$ need be true, and as such, a world at which A may be indeterminate in truth-value.

Given the construction of Logically Ideal Space W_I , it is easy to see that worlds in W_I can only model agents that are logically omniscient. For when we read ‘logical truths’ and ‘logical entailments’ to be solely about propositional logical truths and propositional logical entailments, the construction of W_I allows us to establish (Omni₁) and (Omni₂):

(Omni₁) For any sentences A and B such that A logically entails B , any agent that knows A also knows B .

(Omni₂) For each logical truth T , each agent knows T .

For (Omni₁), suppose A logically entails B , and suppose agent a knows A . By (Know), A is true at all worlds that are epistemically possible for a . Since all worlds in W_I are consistent, all worlds that verify A also verify B . So B is true at all worlds that are epistemically possible for a whenever A is. By (Know), it follows that a knows B whenever a knows A —regardless of how complicated the logical entailment from A to B is. So (Omni₁) holds. For (Omni₂), since each (propositional) logical truth T is entailed by the empty set, each T is true at each $w \in W_I$. So T is true at any w that may remain epistemically possible for any given agent. By (Know), it follows that each agent knows T —regardless of how complicated the logical truth T is. So (Omni₂) holds.

Clearly, (Omni₁) and (Omni₂) are only plausible for ideal, logically omniscient agents with unbounded cognitive capacities.⁸ So to make (Know) plausible for moderately ideal agents that can only engage in limited logical reasoning, we cannot identify worlds in (Know) with worlds in W_I —and hence with entities that are akin to *logically possible worlds*. Rather, with a view to condition (R₁) in (Result), we

⁸Robert Stalnaker would disagree here. For he has argued that the problem of logical omniscience, when properly understood, can be tackled without giving up the standard possible-world framework (see Stalnaker 1984). On Stalnaker’s view, logical omniscience can be explained away using a combined fragmentation and metalinguistic strategy. I believe that there are serious problems with Stalnaker’s strategy, but I will not argue for this claim here. For detailed critiques of Stalnaker, see Bjerring & Tang 2011, Field 2001, chapter 3, Robbins 2004, and Stanley 2010.

must consider worlds that are akin to Hintikka’s *impossible possible worlds*. Following Hintikka, this means admitting in modal space “worlds which look possible and hence must be admissible as epistemic alternatives but which none the less are not logically possible.” (Hintikka 1975, p. 477). Let us call worlds that fail to be logically consistent—in the sense of (Con) above—*impossible worlds*, and let us call modal spaces that contain impossible worlds *logically non-ideal spaces*.

Insofar as we have a good grip on constructions of possible worlds as maximally specific sets of sentences that are logically consistent, we should also have a good grip on the constructions of impossible worlds that I will investigate in what follows. For in these constructions, impossible worlds will simply correspond to maximally specific sets of sentences that fail to be logically consistent.

5 Logically Extreme Space

We need impossible worlds in modal space to model logically non-omniscient agents. In this section, I investigate a model of logically non-ideal space that I call *Logically Extreme Space*. In contrast to Logically Ideal Space, Logically Extreme Space will contain an abundance of impossible worlds that will enable us to model logically non-omniscient agents. Yet, to make (Know) plausible for moderately ideal agents, I argue that we have good reason to investigate models that have a less trivial structure than Logically Extreme Space. In particular, with a view to condition (R₃) in (Result), I argue that we need to investigate models of modal spaces that contain no blatantly impossible worlds.

The construction of Logically Extreme Space only requires two of the ingredients from the construction of Logically Ideal Space. First, I identify worlds with arbitrary sets of sentences in \mathcal{L} . Second, I retain the principles (Truth) and (Falsity) that treat truth and falsity as set-inclusion and set-exclusion respectively. We can then immediately establish that worlds obey:

(Basic Maximality) For all sentences A and worlds w , either A is true at w or A is false at w .

Because $A \in w$ or $A \notin w$, for each sentence A and world w , (Basic Maximality) follows trivially from (Truth) and (Falsity). Let Logically Extreme Space be the class W_E of worlds that satisfy these basic principles.

As above, (Basic Maximality) ensures that worlds can deliver an answer to each question that could in principle be asked about a particular way the world might be. The question “Is A true in the maximally specific hypothesis that world w describes?” will be answered yes if $A \in w$, and no if $A \notin w$. But since worlds in W_E need not be logically consistent, there are plenty of impossible worlds in W_E at which logical truths can be false. More specifically, for *any* sentence A , there is a world $w \in W_E$ such that A is false at w , and for *any* set of sentences $\{A_1, A_2, \dots\}$, there is a world $w \in W_E$ such that the truths at w are exactly A_1, A_2, \dots ⁹

By identifying worlds in (Know) with worlds in W_E , we can ensure that (Know) applies to logically non-omniscient agents. To see this, notice that we can easily invalidate both (Omni₁) and (Omni₂) using worlds in W_E . First, for *any* two sentences A and B such that A logically entails B , there are worlds in W_E that verify A and falsify B . So even if A is true at all worlds that are epistemically possible for an agent, B need not be true at any one of these worlds. By (Know), it follows that no agent need know even the most simple logical consequences of what she already knows. So (Omni₁) no longer holds. Second, we can pick *any* sentence A and find a world $w \in W_E$ that falsifies A . In particular, for *any* logical truth T , there is a $w \in W_E$ such that $T \notin w$. So T need not be true at any w that remains epistemically possible for any agent. By (Know), it follows that no agent need know any logical truth, however obvious. So (Omni₂) no longer holds.

Logically Extreme Space W_E is well suited for modeling the epistemic states of *extremely non-ideal agents* for whom “[n]o inferences, however obvious and useful, need be made from the beliefs”, and for whom “the belief set can include any and all inconsistencies”. (Cherniak 1986, p. 6). W_E can serve as a model of extremely non-ideal agents because of its “explosive anything goes” nature: for “the most absurd situations conceivable”, there is a corresponding world in W_E . (Nolan 1997, p. 544). Everything goes in W_E because there are no non-trivial structural and formal constraints on worlds in this space. Pick ten arbitrary sentences A_1, A_2, \dots, A_{10} , collect them in a set Γ , and you have a world $w \in W_E$ that corresponds to Γ . So arbitrary blatant inconsistencies like {‘It rains’, ‘It does not rain’}, and arbitrary joint inconsistencies like {‘It is summer and it is sunny’, ‘It is not summer’, ‘It is not sunny’}

⁹Structurally, worlds in W_E are akin to the *open worlds* in Priest 2005. Roughly, we can characterize an open world as a world at which any sentence can be evaluated for truth and falsity completely independently of the evaluation of any other sentence at that world; cf. Priest 2005, p. 22.

are among the sentences verified by particular worlds in W_E .

Although this trivial structure is useful for modeling extremely non-ideal agents and for drawing arbitrarily fine-grained distinctions among possibilities, it is also clear that it is of less use for modeling the broad class of moderately ideal agents that are logically competent but not logically omniscient.

Moderately ideal agents always rule out as epistemically possible worlds that verify blatant inconsistencies. So no such worlds should survive in a logically non-ideal space for these agents. For if they did, the structure of modal space would not by itself guarantee that moderately ideal agents always know that blatant inconsistencies are false. Since W_E contains an abundance of blatantly impossible worlds, we hence cannot use Logically Extreme Space to ensure that (Know) is plausible for moderately ideal agents.

More generally, Logically Extreme Space is of little use when it comes to isolating a notion of non-trivial content or meaning that can help shed light on core issues in the philosophy of mind and language. It is standard to use *intensions* to represent content in a world-involving framework. Intensions are functions from worlds of some sort to a truth-value. Let the *extreme intension* of a sentence be a function from worlds in W_E to a truth-value. For *any* two sentences A and B , there are worlds w in W_E such that w verifies A and falsifies B . It follows that the values of the extreme intensions of any two sentences can always come apart. So, if extreme intensions represent content, there is nothing more to the contents of sentences—and hence the contents of knowledge and beliefs—than the sentences themselves. Effectively, that is, the contents of sentences and thoughts correspond trivially to the sets of sentences that contain them, and any non-trivial inferential relations between sentences and thoughts need never be reflected in the corresponding extreme intensions. Thus views that impose minimal logical structure on contents will not find the construction of Logically Extreme Space explanatory useful for a set-theoretical representation of the contents of thought and language.

So there is reason to hold that worlds should obey certain substantive constraints that worlds in Logically Extreme Space do not obey. In particular, a *logically non-trivial* space that we can use to model moderately ideal agents should not contain worlds that verify blatant inconsistencies. Rather, with a view to condition (R₃), a logically non-trivial space should contain no blatantly impossible worlds but only possible and subtly impossible ones.

In response to the problems above, one might point out that there are relevant subspaces in Logically Extreme Space W_E that have all the structure that I am requesting.¹⁰ For instance, one might observe that there *are* worlds in W_E that verify basic logical truths such as $\neg(A \wedge \neg A)$. To ensure that a moderately ideal agent always comes out as knowing that $(A \wedge \neg A)$ is false, we can then simply focus on this class of worlds in W_E to gain the required structure. I agree. But we can still demand information about the structural properties of the relevant subspace. Obviously, it is not useful to be told merely that there *is* a class of worlds in W_E that can do the relevant work that we want logically non-ideal space to do. Rather, we want a method or a procedure for isolating the relevant class of worlds. Yet, if one prefers, one can think of attempts to construct a logically non-trivial space as attempts to formally delineate subspaces in W_E that are suitable for modeling moderately ideal agents.¹¹ But the conclusion remains that the *structural* features of W_E are inadequate for modeling such agents.

So I trust that there is good motivation for (R_3) and hence for being interested in constructions of a logically non-trivial space. Logically Extreme Space has its roles to play, but if we want a less trivial model of a space that can help us model moderately ideal agents, we need to go beyond Logically Extreme Space—or at least appropriately far inside it.

For the purpose of this paper, I will not engage in an elaborate and complicated discussion of (Maximality) and (Basic Maximality)—and hence of condition (R_2) . I am not claiming that people who are sympathetic to the impossible-world approach to non-ideal rationality would want to maintain a maximality principle. But I believe it is worthwhile to separate issues that concern the inconsistency of worlds from issues that concern the maximality of worlds. For these issues are conceptually quite different, and the problems that each principle raises seem quite different to me as well. I will briefly substantiate these claims below, although a full elaboration of these points will require a paper on its own.

First, (Basic Maximality)—and (Maximality) likewise—says that each sentence of

¹⁰Thanks to Jonathan Schaffer for discussion here.

¹¹Alternatively, as mentioned in section 2, one can think of attempts to construct a logically non-trivial space as attempts to formally delineate the class E of worlds in W_E that may remain epistemically possible for moderately ideal agents. For such agents, E cannot contain blatantly impossible worlds. For extremely non-ideal agents, in contrast, E may contain blatantly impossible worlds, and for ideal agents, E can contain only possible worlds.

the language \mathcal{L} is either true or false at each world in modal space. But it is quite compatible with what I have said about \mathcal{L} that it is a rather small or limited language, which does not contain certain atomic sentences, or even whole domains of sentences. Insofar as the exact size of \mathcal{L} can vary—maybe in accordance with what is at issue in a given context—we can leave open the possibility that many sentences do not receive a truth-value at worlds. I take it that one of the main philosophical motivations for going for partial worlds is the idea that there are sentences—or even entire topics—that agents are completely unaware of or that they cannot grasp cognitively. If this is true, we can in part accommodate this motivation for partiality by suitably restricting \mathcal{L} rather than giving up (Maximality) or (Basic Maximality).¹²

Second, (Basic Maximality)—and (Maximality) likewise—is a standard principle in a two-valued world-involving framework. And it is a principle that we can aim to retain even if we admit logically inconsistent worlds in modal space. By retaining a maximality principle, we can not only stay closer to the standard possible-world framework but also bracket specific worries that one might have about non-maximal or partial worlds but not about logically inconsistent worlds. By way of example, consider the standard Kripkean analysis of the \diamond operator as “truth in at least one (accessible) world.” If we admit partial worlds in modal space, it turns out to be complicated to establish this simple analysis of \diamond ; see Correia 1999 and Schotch et al. 1978. But similar complications need not arise with respect to modal spaces that contain logically inconsistent but maximal worlds. In this sense, one might have specific worries towards the use of partial worlds that do not pertain to the use of logically inconsistent but maximal worlds.

Third, (Basic Maximality)—and (Maximality) likewise—allows us to sidestep issues about the notion of indeterminacy. If A is neither true nor false at a world w , then A is indeterminate at w . But what does it mean for a sentence to have an indeterminate truth-value at a world? Do we interpret indeterminacy as a *third* truth-value alongside truth and falsity, or instead as a mere *lack* of a truth-value? I expect that different answers to these questions will lead to different models of modal space with different complications.

If indeterminacy is interpreted as a third truth-value, it seems that agents should

¹²The idea of capturing intuitions about partiality by restricting \mathcal{L} while retaining a maximality principle mirrors Stalnaker’s use of (internal) maximality. According to this notion of maximality, each world settles the truth-value of everything that can be said *within a given context* (Stalnaker 1986).

be able to reason about indeterminacies. If so, it seems reasonable to hold also that minimally rational agents can rule out partial worlds such as w^* at which only $(A \wedge B)$ is true but all other sentences indeterminate. For, on the face of it, if a minimally rational agent knows that a conjunction is true, then she also knows—or at least is in a position to know easily—that the conjuncts are *true* rather than *indeterminate*. This observation raises the question of how agents can rule out partial worlds. On one natural understanding, an agent can rule out a partial world w by either inferring a contradiction from what holds at w , or by inferring that some sentence A , which w says is indeterminate, actually is provably true or false according to what holds at w . If this is the general idea, I believe problems similar to those in section 6 will arise, but I cannot argue for it here (see Bjerring 2010).

Alternatively, if indeterminacy is interpreted as a lack of a truth-value, it may be that agents are not able to reason about indeterminacies. Indeterminacies, so to speak, mirror dark spots in the agent’s epistemic landscape, whether these dark spots are caused by unawareness, lack of conceptual or representational means, or some such phenomena. If we think of indeterminacies in this way, it may also seem reasonable to hold that agents need not rule out partial worlds like w^* . But this depends on the work that we want partial worlds to do. If the resulting partial modal space is to serve as a proper space for minimally rational agents, worlds cannot be *too* partially described. Plausibly, just as we expect minimally rational agents to rule out blatant logical inconsistencies, so we expect them to accept *obvious* or *self-evident* logical truths. But this means that worlds like w^* do not really describe genuine possibilities for minimally rational agents.¹³ For w^* does not verify obvious logical truths such as $(A \rightarrow A)$, $(A \vee \neg A)$, and $\neg(A \wedge \neg A)$. If this is true, worlds that are *too* partially described should be excluded from a partial modal space. But then a host of new questions arise. How partial can partial worlds be? And how do we define procedures for eliminating worlds that are too partially described? Answers to these questions lie well beyond the scope of this paper, but I hope to address them and the partial impossible-world approach to logical omniscience in future work.

In light of this, one may take (Result) to show that inconsistency is not enough for a successful impossible-world framework: non-maximal or partial worlds are needed

¹³Hintikka implies something similar when he writes that partial worlds must be “large enough to stand on their own feet in the sense of being large enough to show that the state of affair in question is really possible.” (Hintikka 1969, p. 26.)

as well. For people who are antecedently fond of both partial and inconsistent worlds, they can take (Result) to provide a further reason for letting a serious maximality principle go.

6 Logically Non-Trivial Space: an Impossibility

At this point we know that we need to steer clear of both Logically Ideal Space and Logically Extreme Space to model moderately ideal agents. That is, we need a logically non-trivial space in which there are impossible worlds but no blatantly impossible worlds. But in this section I prove that there is no such modal space when worlds are maximal.¹⁴ In particular, I will prove that (Incon) holds when worlds are maximal:

(Incon) All logically impossible worlds are blatantly impossible.

If we can show (Incon), we will have shown that blatantly impossible worlds must be admitted in modal space if we want to avoid logical omniscience. In turn, this will establish (Result).

To prove (Incon), I will show that (G-Incon) holds—‘G’ for ‘generic’:

(G-Incon) All maximal, logically inconsistent sets of sentences contain an instance of a LNC-, MP-, or NC-inconsistency:

LNC-inconsistency (law of non-contradiction): $\{A, \neg A\}$.

MP-inconsistency (modus ponens): $\{A, (A \rightarrow B), \neg B\}$.

NC-inconsistency (negated conditional): $\{\neg A, \neg(A \rightarrow B), \neg B\}$, $\{\neg A, \neg(A \rightarrow B), B\}$, $\{A, \neg(A \rightarrow B), B\}$.

As always, sets of sentences are formed over the language \mathcal{L} that has symbols \neg and \rightarrow that play the same inferential roles as classical negation and material implication. Since we can define the other standard connectives in terms of \neg and \rightarrow , logically inconsistent sets such as $\{B, (A \wedge \neg B)\}$ do not make (G-Incon) false since they are treated as shorthand for $\{B, \neg(A \rightarrow B)\}$.

I will prove (G-Incon) by using (Sat) from section 4:

¹⁴We remember that when a world w is maximal, then for all sentences A , either A is true at w or $\neg A$ is true at w . By the account of truth and falsity and the identification of worlds with sets of sentences, this means effectively that for all A , either $A \in w$ or $\neg A \in w$.

(**Sat**) Any set of sentences Γ that satisfies the following two conditions is satisfiable:

- (i) $A \in \Gamma$ iff $\neg A \notin \Gamma$.
- (ii) $(A \rightarrow B) \in \Gamma$ iff $A \notin \Gamma$ or $B \in \Gamma$.

Given (**Sat**), the proof of (**G-Incon**) goes as follows:

Proof Let Δ be any maximal, logically inconsistent set of sentences. By (**Sat**), then Δ will fail to satisfy either (i) or (ii) and hence contain at least one of the following inconsistent pairs or triples of sentences:

Case 1: Δ may be inconsistent because it fails to satisfy (i) of (**Sat**), in which case Δ contains an inconsistent pair of the form $\{A, \neg A\}$. That is, Δ contains an instance of a LNC-inconsistency.

Case 2: Δ may be inconsistent because it fails to satisfy (ii) of (**Sat**), in which case Δ contains either an inconsistent triple of the form $\{A, (A \rightarrow B), \neg B\}$, or an inconsistent triple of the form $\{\neg A, \neg(A \rightarrow B), \neg B\}$, $\{\neg A, \neg(A \rightarrow B), B\}$, or $\{A, \neg(A \rightarrow B), B\}$. That is, Δ contains either an instance of a MP-inconsistency or an instance of a NC-inconsistency.

Derivatively, all maximal, logically inconsistent sets of sentences contain an instance of a LNC-, MP-, or NC-inconsistency. Thus (**G-Incon**) holds. ■

Given (**G-Incon**), we can now establish (**All-Blatant**):

(**All-Blatant**) Each instance of a LNC-, MP-, and NC-inconsistency is a blatant inconsistency.

The proof of (**All-Blatant**) goes as follows:

Proof I show that a contradiction can be inferred from each instance of a LNC-, MP-, and NC-inconsistency by use of at most 1 application of an inference rule.

Case LNC-inconsistencies: Each instance of a LNC-inconsistency is already a contradiction.

Case MP-inconsistencies: A contradiction can be inferred from each instance of a MP-inconsistency by use of 1 application of modus ponens.

Case NC-inconsistencies: A contradiction can be inferred from each instance of a NC-inconsistency by use of 1 application of (R-NC)—(R-NC) is the rule for negated conditionals that permits inferring in 1 step A and $\neg B$ from $\neg(A \rightarrow B)$. So a contradiction can be inferred from each instance of a LNC-, MP-, and NC-inconsistency by use of at most 1 application of the inference rules. It follows by (Blatant) that each instance of a LNC-, MP-, and NC-inconsistency is a blatant inconsistency. Thus (All-Blatant) holds. ■

Worlds correspond to maximal sets of sentences, and a world is blatantly impossible just in case it contains a blatant inconsistency. By (G-Incon), all impossible worlds contain an instance of a LNC-, MP-, or NC-inconsistency. Then, by (All-Blatant), all impossible worlds contain a blatant inconsistency. So we have proved:

(Incon) All logically impossible worlds are blatantly impossible.

Alternatively, we have proved that all impossible worlds can be ruled out using just 1 step of logical reasoning.

Given (Incon), (Result) follows immediately:

(Result) There is no modal space such that:

- (R₁) there are impossible worlds;
- (R₂) there are no partial worlds;
- (R₃) there are no blatantly impossible worlds.

For any modal space W that satisfies (R₂) and (R₃), (Incon) tells us that it cannot also satisfy (R₁): if W contains no blatantly impossible worlds, then it contains *no* impossible worlds at all. As a result, W will collapse to a modal space akin to Logically Ideal Space that is useful only for modeling logically omniscient, ideal agents. So to model logically non-omniscient agents, W must contain blatantly impossible worlds. But as a result, W will collapse to a modal space akin to Logically Extreme Space that is useful only for modeling extremely non-ideal agents. So when worlds are maximal, I conclude that we cannot use an impossible-world framework to ensure that (Know) is plausible for moderately ideal agents.

It is clear that the proof of (G-Incon) and (All-Blatant) depend on the specific inference rules that represent logical reasoning. But for most normal proof and semantic tableaux systems, instances of LNC-, MP-, and NC-inconsistencies can be ruled

out in relatively few steps. This will maintain the force of (Incon) and allows us to conclude that all impossible worlds can be ruled out using just a few steps of logical reasoning.

More generally, however, there is a good pre-theoretical sense in which instances of LNC-, MP, and NC-inconsistencies count as blatantly inconsistent. A blatant inconsistency, on one very natural way of understanding the term, is an inconsistency that any minimally rational agent can easily rule out using solely logical reasoning. Irrespective of how precisely we represent the kind of logical reasoning that such agents can engage in, it remains intuitively clear that they can perform the easy, obvious and feasible logical reasoning that suffices to rule out LNC-, MP-, and NC-inconsistencies. Given (G-Incon), we can hence establish (Result) by use of any definition of blatant inconsistencies that characterizes instances of LNC-, MP, and NC-inconsistencies as blatantly inconsistent. This indicates the generality of (Result).

Presumably, as well, we will be able to establish results similar to (G-Incon) for logics that are stronger than propositional logic. To get such results, the task is to isolate a few kinds of basic inconsistencies in the stronger logic that occur in all inconsistent sets of sentences. When these sets are also maximal, and when the logic in question allows us to move from simple logical truths to complex ones by repeated applications of 1-step inference rules, it seems likely that a result similar to (G-Incon) can be shown. In turn, this would allow us to establish (Result) for worlds that are impossible by the lights of logics stronger than propositional logic.

7 Conclusion

To model moderately ideal agents, we need to find a construction of a logically non-ideal space that can avoid the Charybdis of logical omniscience and the Scylla of “anything goes”.¹⁵ Whereas impossible worlds must be included in modal space to avoid the Charybdis of logical omniscience, blatantly impossible worlds must be excluded to avoid the Scylla of “anything goes”. But when worlds are maximal, (Result) shows that we must accept the Scylla of “anything goes” to avoid the Charybdis

¹⁵Cf. Chalmers 2011, p. 103:

Perhaps the biggest open problem in the study of non-ideal epistemic space is that of finding a construction of non-ideal [worlds] that avoids the Scylla of “anything goes” and the Charybdis of logical omniscience.

of logical omniscience.

Less metaphorically, I have shown in this paper that it is impossible to have an impossible-world framework in which worlds are maximal and in which we can model moderately ideal agents. To avoid this impossibility result, we must give up either (R_1) , (R_2) or (R_3) . Insofar as we want to model logically non-omniscient agents using impossible worlds, we need condition (R_1) . Insofar as we want to construct a modal space that has a less trivial structure than Logically Extreme Space, we need condition (R_3) —or at least something in the vicinity of (R_3) . So, in the current context, the most promising strategy seems to involve investigating logically non-trivial spaces that contain partial worlds. This leaves us with (R_2) . If we give up (R_2) , we allow that partial or non-maximal worlds may be included in a logically non-ideal space. In section 5, I discussed briefly some of the issues that arise if we give up a serious maximality principle. Nothing I have said in this paper rules out a successful construction of a partial modal space, although it does tell us about the direction that such a construction should go. Investigations of partial modal spaces lie beyond the scope of this paper, but I intend to deal with them in future work.

For now, however, I side with Lewis; an impossible-world framework requires a distinction between subtly and blatantly impossible worlds, and I just do not understand how to draw that distinction.

References

- Adams, R. (1974). Theories of Actuality. *Noûs*, 8, 211-231.
- Bjerring, J. C. & Tang, W. H. (2011). Undermining the Stalnakerian Solution to the Problem of Logical Omniscience. Manuscript.
- Bjerring, J. C. (2010). *Non-Ideal Epistemic Spaces*. Ph.D. thesis, Australian National University.
- Bostock, D. (1997). *Intermediate Logic*. Oxford: Clarendon Press.
- Carnap, R. (1947). *Meaning and Necessity, a Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.
- Chalmers, D. J. (2011). The Nature of Epistemic Space. In A. Egan, & B. Weatherson (Eds.), *Epistemic Modality*. New York: Oxford University Press, 60-107.
- Cherniak, C. (1986). *Minimal Rationality*. Cambridge: MIT Press.
- Copi, I. M. (1979). *Symbolic Logic*. 5th edition. New York: Macmillan.
- Correia, F. (1999). Adequacy Results for Some Priorean Modal Propositional Logics. *Notre Dame Journal of Formal Logic*, 40, 236-249.
- Field, H. (2001). *Truth and the Absence of Fact*. Oxford: Oxford University Press.
- Hintikka, J. (1962). *Belief and Knowledge*. Ithaca: Cornell University Press.
- Hintikka, J. (1969). *Models for Modalities*. Dordrecht: D. Reidel Publishing Company.
- Hintikka, J. (1975). Impossible Possible Worlds Vindicated. *Journal of Philosophical Logic*, 4, 475-484.
- Jago, M. (2009). Logical Information and Epistemic Space. *Synthese*, 167, 327-341.
- Jeffrey, R. (1983). *The Logic of Decision*. Chicago: University of Chicago Press.
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell Publishers.

- Lewis, D. (2004). Letters to Priest and Beall. In B. Armour-Garb, J.C. Beall, G. Priest (Eds.), *The Law of Non-Contradiction—New Philosophical Essays* (pp. 176-177). Oxford: Oxford University Press.
- Nolan, D. (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, 38, 535-572.
- Priest, G. (2005). *Towards Non-Being: the Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.
- Robbins, P. (2004). To Structure, or Not to Structure? *Synthese*, 139, 55-80.
- Schotch, P., Jensen, J., Larsen, P., & MacLellan, E. (1978). A Note on Three-Valued Modal Logic. *Notre Dame Journal of Formal Logic*, 19, 63-68.
- Stalnaker, R. (1986). Possible Worlds and Situations. *Journal of Philosophical Logic*, 15, 109-123.
- Stalnaker, R. (1984). *Inquiry*. Cambridge: MIT Press.
- Stanley, J. (2010). “Assertion” and Intentionality. *Philosophical Studies*, 151, 87-113.